Molecular Mechanisms for Mechanostability in Biomolecular Complexes

Constantin Schöler

München 2017

Molecular Mechanisms for Mechanostability in Biomolecular Complexes

Constantin Schöler

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Constantin Schöler
aus Würzburg

München, den 19.01.2017

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachter: Prof. Dr. Joachim O. Rädler
Tag der mündlichen Prüfung: 28.02.2017

For
Klaus Schulten

Zusammenfassung

In der vorliegenden Arbeit wurden die zugrunde liegenden molekularen Mechanismen für mechanische Stabilität von biomolekularen Komplexen untersucht. Als Modellsystem wurde dabei ein Rezeptor-Ligand-Komplex verwendet, der für die stabile jedoch reversible Bindung von bakteriellen Zellen an Zellulosefasern verantwortlich ist. Die Möglichkeit, solche Bindungen aufzubauen, ist in der Natur von essentieller Bedeutung. Mechanische Stabilität ist vonnöten um selbst unter widrigen Bedingungen intermolekulare Bindungen aufrechtzuerhalten. Gleichermaßen müssen solche Bindungen reversibel sein, um es Mikroorganismen zu ermöglichen, sich dynamisch an veränderte externe Bedingungen anzupassen. Die Physik solcher intermolekularer Bindungen beruht auf der Kombination einiger weniger fundamentaler Wechselwirkungen (z.B. Elektrostatik oder van-der-Waals Kräfte) mit einer Vielzahl molekularer Konformationen. Im Rahmen dieser Arbeit wurden Einzelmolekülkraftspektroskopieexperimente sowie Molekulardynamiksimulationen (in Zusammenarbeit mit der Gruppe von Prof. Klaus Schulten an der University of Illinois) durchgeführt um eben diese Physik besser zu verstehen.

Mittels Einzelmolekülkraftspektroskopie wurde zunächst die mechanische Stabilität des Modellsystems bei Belastung in seiner nativen Geometrie untersucht. Die gemessenen Kräfte fielen dabei in einen Kraftbereich, der dem von kovalenten Bindungen nahe kommt. Weiter wurde durch Molekulardynamiksimulationen festgestellt, dass sich die intermolekulare Kontaktfläche der beiden Bindungspartner durch Umordnung der Aminosäurenseitenketten an der Bindungsstelle unter Last vergrößert, und damit die Interaktion verstärkt wird. Im Folgenden wurde sowohl experimentell als auch mittels Simulationen die Abhängigkeit der mechanischen Stabilität des Systems von der Belastungsgeometrie untersucht. Bei nicht nativer Belastung trat dabei ein zusätzlicher Dissoziationspfad auf, bei dem wesentlich geringere Abrisskräfte gemessen wurden. Um die dafür verantwortlichen Mechanismen besser zu verstehen, wurde ein Verfahren zur Visualisierung der Pfade entwickelt, entlang denen sich externe Kräfte durch einen molekularen Komplex ausbreiten. Hierzu wurde eine Kombination von thermodynamischer Fluktuationstheorie mit netzwerkbasierter Korrelationsanalyse verwendet. Mit diesem Verfahren konnte gezeigt werden, dass sich im Falle erhöhter mechanischer Stabilität die Kraft an der Bindungsstelle lokal auf Pfaden senkrecht zur Belastungsachse ausbreitet und somit einen vergleichsweise kleinen Energiebetrag für die Dissoziation leisten kann.

In einem gesonderten Projekt wurde eine Analysemethode zur Korrektur experimenteller Kraftspektroskopiedatensätze entwickelt. Experimentell werden oftmals Rezeptor-Ligand-Komplexe als spezifischer Zuganker zur Entfaltung von Proteindomänen eingesetzt. Falls die Wahrscheinlichkeitsverteilungen der Entfaltungs- bzw. Abrisskräfte dabei überlappen, können statistische Artefakte auftreten. Die vorgestellte Methode ermöglicht es, die experimentell gewonnenen, verfälschten Verteilungen mathematisch zu entfalten und damit unverfälschte Parameter der zugrundeliegenden Energielandschaften zu extrahieren. Dabei ist die Methode unabhängig vom verwendeten Modell zur Beschreibung der Entfaltungs- bzw. Abrissverteilungen.

Preface

A multitude of biological processes rely on the formation of stable, yet reversible intermolecular bonds. The physics of such bonds are governed by a combination of elementary interactions, such as electrostatics or van-der-Waals interactions, and a multitude of molecular conformations. A particular interesting family of intermolecular bonds are those that need to resist mechanical forces in their native function. This occurs in various cell adhesion processes like leukocyte tethering to vascular surfaces [1], platelet binding to von Willebrand factor at sites of vascular injuries [2, 3], or bacterial cell adhesion to cellulosic food sources [4, 5]. These bonds need to be reversible to allow for dynamic assembly and disassembly of molecular structures to adapt to changing environmental conditions. Additionally, they need to be able to resist mechanical forces that can, for example, be exerted by shear flow gradients in turbulent environments. The former is reflected in common place biochemical affinities (*i.e.*, moderate dissociation rates at equilibrium), while the latter is captured by elevated dissociation forces along selected reaction coordinates. The main objective of this thesis was to understand the molecular mechanisms responsible for mechanical stability in biomolecular complexes at the single molecule level. While bulk methods relying on ensemble averaging provide information on the overall dynamics of intermolecular bonds, single molecule manipulation techniques provide unparalleled insight into the free energy landscapes governing unfolding of protein domains or the dissociation of molecular complexes. To this end, a receptor ligand complex involved in the adhesion of the ruminal bacterium *Ruminococcus flavefaciens* to cellulose fibers was used as model system. The complex is part of the bacterium's cellulosome [6], an intricate extracellular machinery designed to efficiently degrade cellulose enzymatically. The assembly, disassembly, and anchoring of cellulosomes is governed by non-covalent, highly specific cohesin dockerin interactions. In its native function, the type III cohesin dockerin system investigated here anchors a bacterial host cell to its cellulosic food sources and is expected to be subjected to mechanical forces *in vivo*. The mechanics of this system were assessed through a combination of single molecule force spectroscopy experiments and steered molecular dynamics simulations (performed in collaboration with the group of Prof. Klaus Schulten at the University of Illinois).

Chapters 1 and 2 serve as an introduction to the techniques used in this work. Chapter 1, which was based on a recently published review article, provides a detailed introduction to the field of single molecule force spectroscopy, including surface immobilization strategies and theoretical background. The basic concepts of molecular dynamics simulations and special analysis procedures thereof are outlined in Chapter 2.

The main results of this thesis are presented in Chapters 3-5. The results part of this thesis is self contained, *i.e.*, every Chapter is accompanied by its own materials and methods section and should be accessible without knowledge of the prior chapters. The mechanical stability of the model system in its native geometry was investigated both experimentally and through simulations in Chapter 3. Experimentally, it was found that the system resists mechanical forces that fall in a range approaching the stability of covalent bonds. Simulations then revealed that the origin of this exceptional mechanostability is an increase in the intermolecular contact

area achieved through rearrangement of amino acid side chains at the binding interface. The study presented in Chapter 4 sought to then identify how an external force is propagated through a molecular structure. To this end, the dependence of the pulling geometry on the mechanical stability of our model system was tested. Therein, it was found that, when stressed in a non-native geometry, an additional unbinding pathway emerged, that was characterized by much lower dissociation forces. Moreover, a novel analysis technique for molecular dynamics trajectories was developed that allows to visualize force propagation pathways. This was achieved through a combination of thermodynamic fluctuation theory and network based correlation analysis. This approach revealed that in cases of elevated mechanical stability the system is able to locally direct the external force across the binding interface at angles with large normal components to the unbinding axis. This results in a lower energy contribution to the dissociation process when compared to a propagation along the unbinding axis.

In Chapter 5, an analysis protocol for experimental force spectroscopy data to correct for a statistical biasing effect was developed. This effect may occur when receptor ligand complexes are used as specific handles to study the unfolding of protein domains. More specifically, biasing occurs if the distributions of forces associated with domain unfolding and receptor ligand dissociation exhibit a finite overlap. In this case, the experimentally observed distributions need to be deconvolved to extract correct parameters of the underlying free energy landscapes through relevant theoretical models. The analysis protocol presented in Chapter 5 is independent of the theoretical model used to describe the distributions of unfolding and unbinding forces.

The results presented in this thesis were published in form of four peer-reviewed publications:

1. Wolfgang Ott*, Markus A. Jobst*, Constantin Schoeler*, Hermann E. Gaub, and Michael A. Nash, **"Single-molecule force spectroscopy on polyproteins and receptor-ligand complexes: the current toolbox"**, *Journal of Structural Biology* (2016), DOI: 10.1016/j.jsb.2016.02.011.
   *contributed equally to this work

2. Constantin Schoeler*, Klara H. Malinowska*, Rafael C. Bernardi, Lukas F. Milles, Markus A. Jobst, Ellis Durner, Wolfgang Ott, Daniel B. Fried, Edward A. Bayer, Klaus Schulten, Hermann E. Gaub, and Michael A. Nash, **"Ultrastable cellulosome-adhesion complex tightens under load"**, *Nature Communications* 5, 5635 (2014), DOI:10.1038/ncomms6635
   *contributed equally to this work

3. Constantin Schoeler*, Rafael C. Bernardi*, Klara H. Malinowska, Ellis Durner, Wolfgang Ott, Edward A. Bayer, Klaus Schulten, Michael A. Nash, and Hermann E. Gaub, **"Mapping Mechanical Force Propagation through Biomolecular Complexes"**, *Nano Letters* 15, 7370-7376 (2015), DOI: 10.1021/acs.nanolett.5b02727
   *contributed equally to this work

4. Constantin Schoeler, Tobias Verdorfer, Hermann E. Gaub, and Michael A. Nash, **"Biasing effects of receptor-ligand complexes on protein-unfolding statistics"**, *Physical Review E* 94, 042412 (2016), DOI: 10.1103/PhysRevE.94.042412

I have further contributed to two peer-reviewed publications, which have not been included as main results of this thesis:

5. Markus A. Jobst*, Constantin Schoeler*, Klara H. Malinowska, and Michael A. Nash, **"Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy"**, *Journal of Visualized Experiments* 82, e50950 (2013), DOI: 10.3791/50950
*contributed equally to this work

6. Markus A. Jobst, Lukas F. Milles, Constantin Schoeler, Wolfgang Ott, Daniel B. Fried, Edward A. Bayer, Hermann E. Gaub, and Michael A. Nash, **"Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy"**, *eLife* 4, 10319 (2015), DOI: 10.7554/eLife.10319

Contents

Part I

Introduction

Chapter 1

AFM based single molecule force spectroscopy

Summary

Single-molecule force spectroscopy sheds light onto the free energy landscapes governing protein folding and molecular recognition. Since only a single molecule or single molecular complex is probed at any given point in time, the technique is capable of identifying low-probability conformations within a large ensemble of possibilities. It furthermore allows choosing certain unbinding pathways through careful selection of the anchor points at which the force acts on the protein or molecular complex. This introduction, partly published as a review article, focuses on recent innovations in construct design, site-specific bioconjugation, measurement techniques, instrumental advances, and data analysis methods for improving workflow, throughput, and data yield of atomic force microscopy based single molecule force spectroscopy experiments. Current trends that are highlighted include customized fingerprint domains, peptide tags for site-specific covalent surface attachment, and polyproteins that are formed through mechanostable receptor-ligand interactions. Recent methods to improve measurement stability, signal-to-noise ratio, and force precision are presented, and theoretical considerations, analysis methods, and algorithms for analyzing large numbers of force-extension curves are further discussed. The various innovations identified here will serve as a starting point to researchers in the field looking for opportunities to push the limits of the technique further.

1.1 Introduction

The atomic force microscope [8] was initially developed as an imaging technique in air or low temperature and ultra high vacuum conditions. Shortly after, measurements in biologically relevant conditions became possible when the AFM was combined with fluid cells and optical lever detection, *i.e.* the amplification of cantilever motion by a detection laser beam (Fig.

Parts of this chapter were published by Ott *et al.* [7] (Schoeler, C shared first author) in the *Journal of Structural Biology* and adapted with permission from Elsevier Inc.

1.1A) [9]. The early 1990s then saw an explosion of the bio-AFM field, which opened the door to high-resolution imaging of proteins and cell surfaces under near-native conditions [10–12]. Shortly thereafter came the realization that individual proteins and DNA molecules, or single receptor-ligand complexes, could be probed with the help of nano- to microscale force transducers (*e.g.*, cantilevers, optically trapped beads, magnetically trapped beads) [13–18]. It was furthermore discovered that natural polyproteins (*e.g.*, Titin) with repetitive multi-domain structures provided regularly repeating saw-tooth like features in force extension data [19]. Artificial (*i.e.*, recombinant) polyproteins quickly came into fashion as internal molecular controls for investigating mechanical properties of protein domains of interest. Since then, engineering of polyproteins has provided a wealth of information about mechanostable motifs in protein folds [20–22], directional dependence of protein mechanostability [23–26], and modulation of mechanostability by molecular recognition [27].

Today, force spectroscopy and bio-AFM in general are well established as standard tools in the nanobiosciences, and are regularly used for investigating cell adhesion and cell surface properties [28–32], interrogating membrane proteins [33–36], and measuring mechanical properties of proteins [37–41], polysaccharides [42, 43] and DNA [44]. Recent studies have already begun to characterize membrane proteins in vivo by probing their response to external forces on native living cells [45, 46]. There are a number of review articles that thoroughly cover the field from the early years [47–58].

Despite the high level of interest and well-developed method of AFM-SMFS (single molecule force spectroscopy), there have remained several limitations to the technique that prevent researchers from fully taking advantage of mechano-phenotyping of molecules and cell surfaces. Specifically, low experimental throughput and low yield of interpretable single-molecule interaction curves have both hampered the widespread adoption of the method, and its application for studying a large number of proteins. The purpose of this introduction is to highlight recent developments in bioconjugate chemistry, instrumentation, and data processing/algorithms which aim at improving the design process, yield, measurement quality and throughput of AFM-SMFS experiments.

1.2 Unfolding fingerprints

To avoid multiple interactions and unspecific adhesion events and ultimately favor single molecule interactions, low surface immobilization densities are usually applied in AFM-SMFS. Consequently, in typical experiments, many thousand force-extension curves are recorded, but only a fraction of these curves contain useable data that describe the behavior of a single molecule. Typically, the majority of curves ($\sim 80-99\,\%$), by design, contain no interaction, while some contain a multiplicity of interactions that are difficult to interpret, or unspecific adhesion events as measurement artifacts. The experimenter is left searching for a needle in a haystack, looking for single-molecule interactions among a vast excess of uninterpretable force-extension curves. In order to filter the data efficiently, the SMFS community has identified a broad range of proteins that can be used as specific identifiers in unfolding traces. We refer to these domains as ”fingerprints” because they provide a unique unfolding step or ”contour-length increment” of defined length that can be used as a filter during data processing. These fingerprint domains are typically globular protein domains with unfolding forces and length increments varying across a large range. This ability to choose the length increments and unfolding forces of the

fingerprint domains has enabled the design of custom fusion proteins with well-controlled unfolding behaviors. Recent surveys of mechanical properties of different protein domains are provided by Sulkowska and Cieplak [59] and Hoffmann and Dougan [49].

1.3 Receptor ligand SMFS

Protein-protein and protein-small molecule interactions have also been widely analyzed with SMFS. Reports of receptor-ligand SMFS include measurements on biotin-avidin [15, 16, 60–63], antigen-antibody interactions [64–66] along with several other protein-protein or small molecule interactions [50, 67, 68].

One limitation in the standard method of receptor-ligand SMFS is that the signal lacks single-molecule specificity. Depending on the proteins involved and the experimental conditions (*i.e.*, blocking/passivation steps), and since typically no fingerprint molecules are used, it can be difficult to differentiate non-specific interactions from specific protein-protein recognition. A second limitation of many receptor-ligand SMFS experiments is that pulling geometry is not strictly controlled. While in a standard polyprotein experiment, the pulling is applied strictly between the N- and C-termini of each domain, coupling of receptors and ligands to AFM tips and substrates is often done through amide linkages formed between amine groups on the proteins and activated NHS-ester groups on the surface or cantilever. This implicates a diversity of pulling geometries which are not strictly controlled, resulting in rupture force distributions that are smeared out or otherwise distorted.

1.4 Receptor ligand SMFS with fingerprints

Our group has worked on improving the technique for receptor-ligand SMFS out of sheer necessity (Fig. 1.1). We were interested in studying a family of receptor-ligand proteins (*i.e.*, Cohesin-Dockerin, Coh-Doc) involved in carbohydrate recognition and degradation by anaerobic bacteria [5, 69–73]. These protein receptor-ligand complexes are responsible for building up large extracellular networks of structural scaffold proteins and enzymes. They are linked into these structural networks in well-defined and known orientations (*e.g.*, N-terminal or C-terminal anchoring points). It is important to note that when pulling apart a receptor-ligand complex consisting of two proteins, there are four possible terminal pulling configurations (*i.e.*, N-N', N-C', C-N', C-C') (Fig. 1.1B). Many of the Coh-Doc complexes we are interested in possess a clear "physiological" pulling configuration found in nature, and "non-physiological" or "non-native" configurations. To understand their natural mechanical adaptations giving rise to their remarkable assembly strategy, we sought to characterize the mechanical stability of these receptor-ligand complexes in both their native and non-native loading configurations. We found a way to ensure specific interactions by basically combining two previously separate modes of AFM-SMFS (*i.e.*, on polyproteins and receptor-ligand complexes). We fused the Coh and Doc domains separately to different fingerprint domains, and recombinantly produced each construct as a single fusion protein. The fingerprints serve two purposes:

1. they provide site-specific attachment sites through engineered cysteine residues or peptide ligation tags (see below) to strictly control loading geometry;

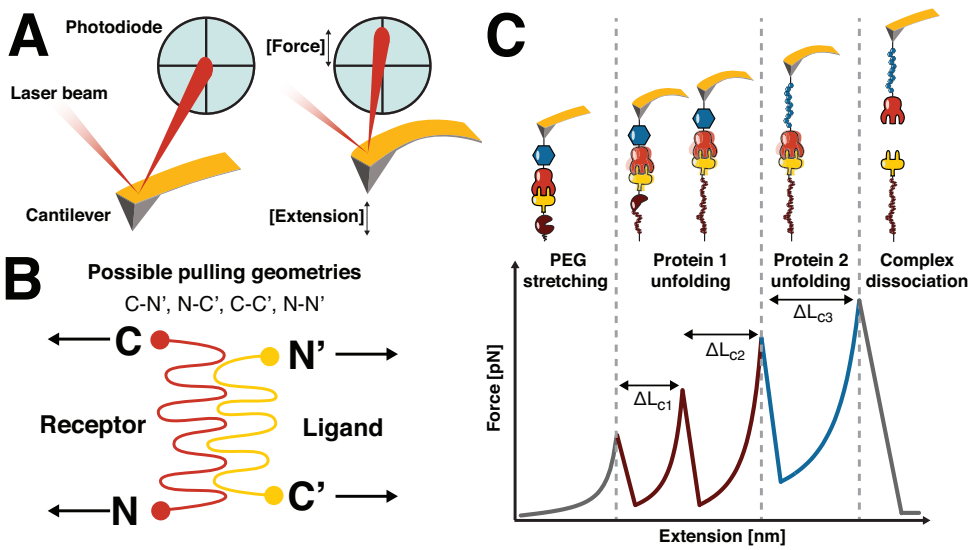


Figure 1.1: Configuration for performing receptor-ligand SMFS with [poly]protein fingerprints. (A) Schematic of the measurement setup. The change of force is detected via the differential signal of the laser beam deflection on a quadrant photodiode. (B) For a protein complex consisting of two domains, 4 terminal pulling configurations are possible (N-N', N-C', C-N', C-C'). (C) Fingerprints (brown and blue) are site-specifically and covalently attached to the cantilever and surface. Receptor (orange) and ligand (yellow) form a stable receptor-ligand complex. Note that the fingerprints can be individual sub-domains, or repetitive polyproteins in their own right. Shown is a typical force-extension trace showing unfolding of the fingerprints, followed by rupture of the receptor ligand complex. In order to observe unfolding of the fingerprints in sufficient numbers, their most probable unfolding force should lie well below the most probable rupture force of the complex for the given loading rate.
Reprinted from Ott *et al.* [7] with permission from Elsevier. Copyright ©2016 Elsevier Inc.

2. they provide predetermined increments in contour length which allows us to filter the datasets for specific single-molecule interactions [5, 69–73].

This configuration yields several advantages: We now have the ability to study mechanical stability of receptor-ligand pairs and unfolding of individual domains (*i.e.*, the fingerprints) in a single-experiment with high yield and specificity, eliminating measurement artifacts. We also have a systematic and straightforward way to probe effects of pulling geometry on receptor-ligand unbinding, and to compare native and non-native pulling configurations. The gene design (*i.e.*, N- or C-terminal fingerprint domains) directly reflects the conformation to be investigated. Furthermore, a specific protein domain of interest can now easily be fused to a mechanostable Coh-Doc receptor-ligand pair for characterization. Depending on the expected domain unfolding forces, an appropriately fitting protein receptor-ligand pair can be chosen from a wide range of well-characterized molecules (Table 1.1). We note that this table does not include every receptor-ligand probed by AFM. For an extensive list of receptor-ligands that were explored with AFM, see Lee *et al.* [50]. Currently, the mechanically most stable receptor-ligand pair is a Coh-Doc type III complex derived from the ruminal bacterium *R. flavefaciens*, with loading-rate dependent rupture forces between 600 to 800 pN [5, 72]. Another interaction in a similar force range range is the trimeric titin-telethonin complex described by Bertz *et al.* [74].

When designing such combined protein unfolding - receptor ligand dissociation experiments, special care has to be taken when selecting the appropriate fingerprint domains:

- In AFM single molecule force spectroscopy experiments, the cantilever sided fingerprint should be able to refold on the timescale of a pulling cycle or the fingerprint will be lost due to a limited number of molecular anchors on the cantilever.

- The distributions of fingerprint unfolding forces should fall in a force range such that receptor ligand dissociation occurs only after all fingerprints have unfolded, *i.e.* there is no significant overlap between the distributions of forces associated with fingerprint unfolding and complex rupture.

If the latter criterion is not fulfilled, a successful experimental pulling cycle, *i.e.* a single receptor ligand connection is established and mechanically dissociated, can have multiple possible outcomes. For a receptor ligand system with one fingerprint, this situation is shown in Fig. 1.2. In this scenario, the complex can either dissociate after fingerprint unfolding (Fig. 1.2, upper path) or with the fingerprint intact (Fig. 1.2, lower path). Since fingerprint unfolding is imposed as a selection criterion during data analysis, all events of the latter type are discarded. This results in a downwards shift of the distribution of fingerprint unfolding forces, and - depending on the pulling protocol - an upwards shift of the observed complex rupture forces. In other words, the experimentally observed distributions for fingerprint unfolding and complex rupture are mutually biased. This effect has first been used qualitatively by Jobst *et al.* [69] (see Appendix E) to unambiguously identify two binding modes with different mechanical stability in a type I cohesin dockerin system. In Chapter 5, a theoretical framework to extract unbiased parameters of the underlying free energy landscapes governing fingerprint unfolding or complex rupture from biased experimental data is presented and validated using a Monte Carlo approach. The calculations are independent of the specific model used to theoretically

Table 1.1: Overview of selected receptor-ligand pairs usable as specific handles for protein-based SMFS experiments. Rupture forces depend on immobilization sites for surface conjugation. Note that rupture forces can also vary depending on probe spring constants and loading rates. Abbreviations: NHS: N-Hydroxysuccinimide; PEG: poly(ethylene glycol); Mal: Maleimide; Cys: Cysteine; CoA: Coenzyme A; SFP: 4-phosphopantetheinyl transferase; ybbR-tag: Peptide sequence DSLEFIASKLA; LF: Low force unbinding path; HF: High force unbinding path. For the column 'immobilization method', the terminology X (Y) Z means: molecule X is attached to Z mediated by enzyme Y.
Adapted from Ott *et al.* [7] with permission from Elsevier.

Protein handles	Sizes [kDa]	Dissociation force [pN]	Immobilization method	Attachment (N/C)	References
Coh: Doc I	15.4/8.3	122.5 ± 18.5	NHS-PEG5000-Mal/Cys	C:C	Stahl *et al.* [73]
Coh: Doc III	21.6/26.2	606 ± 54	NHS-PEG5000-Mal/Cys	N:C	Schoeler *et al.* [72]
		111 ± 30 (LF)	NHS-PEG5000-Mal/CoA (SFP) ybbR	C:C	
		597 ± 67 (HF)	NHS-PEG5000-Mal/CoA (SFP) ybbR		
NiNTA: HIS6	0.2/0.8	153 ± 57	Gold-Cys	n.a.	Verbelen *et al.* [75]
Avidin: biotin	66 − 69/0.2	160 ± 20	Biotinylated BSA	n.a.	Florin *et al.* [60]
StrepTagII: streptavidin	1.1/52.8	253 ± 20	BSA/NHS-biotin	n.a.	Wong *et al.* [76]
Streptavidin: biotin	52.8/0.2	200	Biotinylated BSA	n.a.	Rico and Moy [62]
Calmodulin: CBP	16.7/1.1	16.5 ± 1.8	Pulldown via NI-NTA	n.a.	Junker and Rief [77]
StepTagII: mono-streptactin	1.1/58.4	116	NHS-PEG5000-Mal/Cys	C:C	Baumann *et al.* [78]
		46	NHS-PEG5000-Mal/CoA (SFP) ybbR	N:C	
AntiGCN4 sFv: GCN4(7P14P)	26.7/4.0	70	NHS-PEG5000-Mal/Cys	N:C	Morfill *et al.* [65]
Anti-digoxigenin: digoxigenin	170/0.4	40	NHS-PEG6000	n.a.	Neuert *et al.* [79]

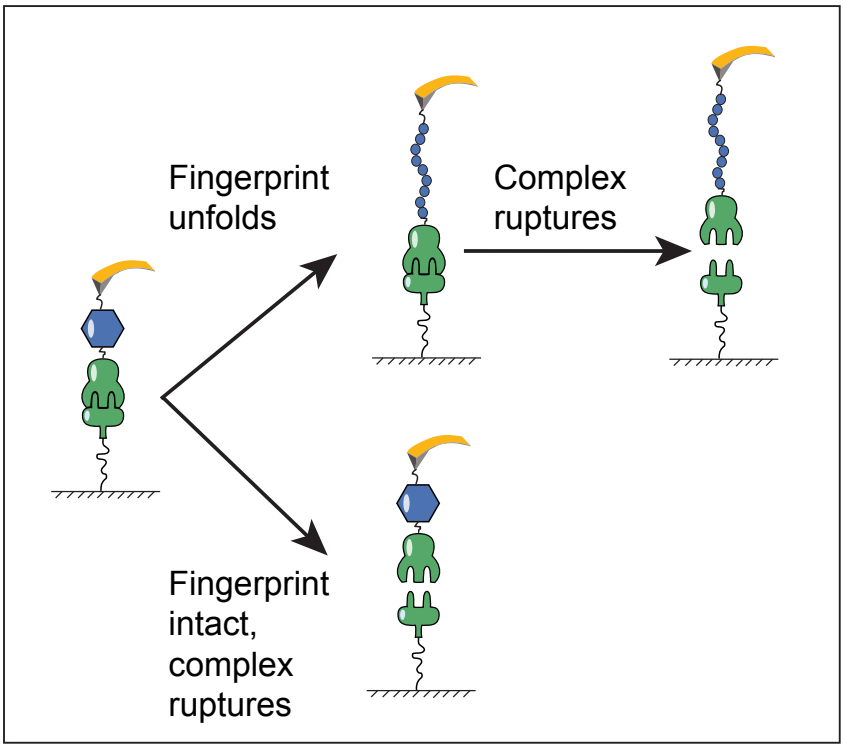


Figure 1.2: Schematic of possible outcomes for receptor ligand dissociation with a single fingerprint and overlapping distributions for unfolding and complex dissociation. Analyzable data show fingerprint unfolding followed by complex rupture (upper path). Due to the overlapping distributions for unfolding and rupture, complex rupture with intact fingerprint is also possible (lower path).
Adapted from Schoeler *et al.* [80] with permission from the American Physical Society.

describe the distributions of forces associated with fingerprint domain unfolding and receptor ligand dissociation.

1.5 Site-specific bioconjugation

Many polyprotein experiments rely on non-specific adsorption of polyproteins onto surfaces (*e.g.*, mica, gold). Receptor-ligand AFM-SMFS, however, requires covalent immobilization of the two binding partners to the cantilever and surface in order to avoid clogging of the molecules on the cantilever tip. Site-specific (*i.e.*, residue specific) conjugation methods provide strict control over the pulling geometry and will result in higher accuracy, precision and reproducibility, compared to conjugation methods resulting in a multiplicity of possible linkage sites (*e.g.*, amine-targeting). Fig. 1.3 provides an overview of established surface chemistry strategies. The two linkage strategies applied in this work are listed below.

1.5.1 Cysteines

Cysteines are relatively rare in proteins, making them attractive as a point mutation residue. The thiol side chain of cysteine is nucleophilic, and will spontaneously react with maleimide leaving groups at neutral pH. It can be used to site-specifically attach proteins to poly ethylene glycol (PEG) coated surfaces for receptor-ligand AFM-SMFS. Alternatively, engineered cysteines can also be used as oligomerization sites to create disulfide-linked polyproteins, as was done for green fluorescent protein (GFP) [81]. However, cysteine/thiol-based protein conjugation has a number of drawbacks, including the tendency of cysteine-modified proteins to multimerize and ultimately aggregate over time, and incompatibility with proteins displaying cysteines on their surfaces in their wild-type form. Hence several other conjugation strategies were developed to overcome this challenge. Most of the newer techniques rely on N- or C-terminal attachment sites because the length of the requisite peptide tags or fusion domains makes inclusion into internal sites of a folded protein domain more challenging.

1.5.2 ybbR/SFP

The ybbR-Tag is an 11 amino acid protein sequence that is enzymatically linked to coenzyme A (CoA) by 4'-Phosphopantetheinyl Transferase (Sfp) enzyme [82–84]. The ybbR-tag has been shown to be N- and C-terminally active. The tag can also be internally incorporated, if permitted by the structure of the protein, however, proper folding is not guaranteed and must be evaluated on a case-by-case basis.

1.5.3 Surface chemistry

The standard protein immobilization approach used for the experimental part of this thesis follows the protocol described by Zimmermann *et al.* [85]. Amino-silanized glass slides and cantilevers are functionalized with a hetero-bifunctional poly(ethylene glycol) (PEG) polymer with an N-hydroxysuccinimide group and a maleimide group at opposing ends. PEG coating provides a passivated surface that resists nonspecific protein adhesion, reducing background and artifacts during measurement. The entropic elasticity behavior of PEG (*i.e.*, persistence length) is similar, although not equal to that of protein backbones, making it a suitable choice for surface conjugation in AFM-SMFS, without interfering too strongly with data interpretation. The maleimide group can then either be modified with CoA containing an inherent thiol group to proceed with ybbR/SFP chemistry, or alternatively directly be reacted with a protein

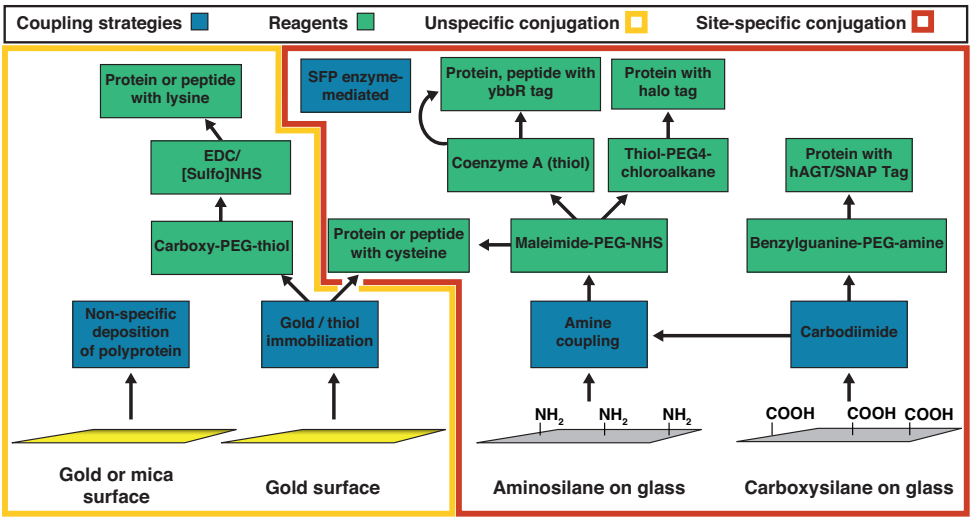


Figure 1.3: Surface chemistry and bioconjugation strategies for single-molecule force spectroscopy. The diagram is by no means exhaustive and is roughly divided into site-specific conjugation methods that provide a single anchoring point for proteins to surfaces/cantilevers (right), and unspecific conjugation methods that provide a heterogeneity of loading configurations (*i.e.*, a multiplicity of pulling points) (left).
Reprinted from Ott *et al.* [7] with permission from Elsevier. Copyright ©2016 Elsevier Inc.

domain displaying a reduced cysteine residue. The PEG incubation can be modified or extended depending on the requirements of the linker and the end group.

1.6 Advances in measurement techniques

Current advances in measurement resolution, instrument stability and accessible dynamic ranges open up new opportunities for measurements of biomolecules. Here we highlight recent innovations aimed at improving quality and precision of AFM-SMFS measurements.

1.6.1 Time resolution in force spectroscopy

In general, the timescales relevant for protein un-/folding and the corresponding timescale for thermally induced crossing of energy barriers are not fully detectable by common SMFS techniques, which typically resolve slower than 50 µs. Early on, the importance of developing high-speed AFM imaging and force spectroscopy through miniaturization of cantilevers with high resonance frequencies and low viscous drag coefficients was appreciated [86, 87]. Nonetheless, only recent studies were able to overcome timescale limitations to observe, for example, extraordinarily slow protein misfolding transitions (~ 0.5 ms) using optical tweezers [88]. Furthermore, advanced statistical methods extended optical tweezers SMFS time resolution to the ~ 10 µs range [89], and optimization of AFM cantilevers for SMFS has pushed the limit toward resolution on the microsecond timescale [90]. These developments allow experimentally accessible ranges to approach the lower limits of fast folding transition dynamics [91, 92], resolving short-lived intermediate states and yielding important insights into other fast conformational dynamics.

1.6.2 Bridging the timescale gap to steered molecular dynamics simulations

Recently, experimental measurements were brought into proximity [72, 93, 94] or even overlap [95] with all atom steered molecular dynamics (SMD) simulations. Depending on the size and thus complexity of the simulated system, it has so far been possible to achieve SMD simulation timescales in the nanosecond to mid-microsecond range [96–98]. Rico *et al.* [99]developed a high speed force spectroscopy AFM based on an Ando-type high speed imaging AFM [100], with a high resonance frequency (600 kHz) miniature multilayer piezoelectric actuator (calibrated before each experiment and run in open loop mode), and a short cantilever with a high resonance frequency (550 kHz in liquid), and low viscous damping. This system was used to record protein unfolding data at extremely high speeds. To reduce hydrodynamic drag, the sample surface was tilted against the direction of the movement. With these improvements and data acquisition in the megahertz range, they were able to record meaningful and interpretable data at pulling speeds of up to 4000 $\mu m\,s^{-1}$, which is about 2 to 3 orders of magnitude faster than conventional methods and starts overlapping with the range of SMD simulations [99]. Despite these successes, care must be taken because underdamped or "ringing" cantilevers like the ones used here are not in agreement with the basic assumptions of the traditional SMFS framework, but can be overcome by custom cantilever optimization procedures at the cost of time resolution [90].

1.6.3 Long-term stability and force precision

Sophisticated measurements of complex biological systems or single molecules often require extraordinarily stable low-drift instruments, capable of continuous long-term data acquisition to gain sufficient and reliable statistics. Active stabilization techniques were developed to enable routine long-term stability and angstrom scale precision at room temperature for optical trap setups: differential sample position was measured and regulated with two independently stabilized and megahertz modulated lasers, backscattered on sample and probe, and recorded separately on a single photodiode using lock-in amplifiers [101]. This method is deemed applicable to surface-based and dual-beam optical traps, magnetic tweezers, AFM setups and optical microscopy, including super-resolution techniques.

AFM cantilever long-term stability and force precision can be increased even further by partially removing the reflective gold coating from the cantilever to dramatically reduce cantilever bending caused by the bimetallic effect [102]. Stability and precision improvements, which still retain high measurement bandwidths, enable and improve on picoscale force and sub-nanoscale motion measurements of molecular properties and dynamics in various biological systems. These may include groundbreaking investigations like the observation of single RNA polymerase base pair stepping [103, 104], base pair unwinding of helicases [105] and prion misfolding pathways [88, 106]. More details on long-term stability measurements and force precision are covered in a recent review of Edwards and Perkins [107].

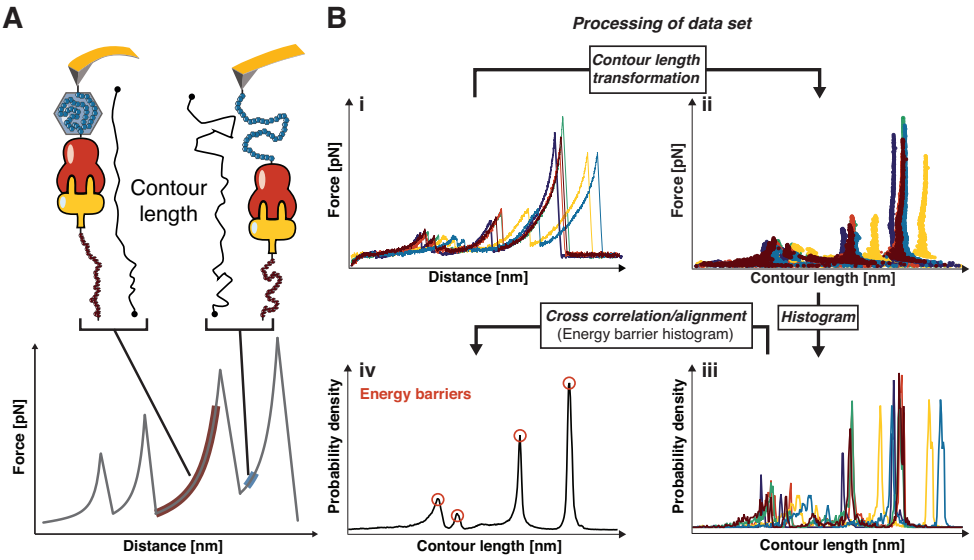


Figure 1.4: Assembly of contour length histograms for screening AFM-SMFS datasets. (A) Force-extension traces are transformed into contour length space using an appropriate polymer elasticity model. The choice of the model depends on the force range in focus. Typical elasticity models for this purpose include the worm-like chain (WLC) at low force (< 150 pN), and the quantum-mechanically corrected freely rotating chain model (QM-FRC) at high force (< 150 pN). (B) Following transformation, the data are plotted in force-contour length space (ii). Force and contour length thresholds are applied and the data are histogrammed (projected onto contour length axis) with an appropriate bin width, *i.e.* nanometer scale, to obtain the diagram in (iii). Each trace analyzed this way can be searched for a specific contour length increment (distance between two peaks in the probability density vs. contour length plot) corresponding to one of the fingerprints. To obtain a master histogram describing all the observed increments in a dataset, individual histograms reflecting a specific unfolding pathway are aligned by cross-correlation and offsetting along the contour-length axis (iv).
Reprinted from Ott *et al.* [7] with permission from Elsevier. Copyright ©2016 Elsevier Inc.

1.7 Theory and data analysis

1.7.1 The data analysis problem

Technical advances that greatly increase the throughput of AFM-SMFS measurements have made automated data analysis protocols an essential requirement. In practice, researchers face the problem of extracting meaningful single molecule signal from large datasets that contain an abundance of unusable data. The use of well-defined fingerprint domains with known unfolding patterns facilitates this procedure greatly. To avoid tedious and time-consuming manual sorting of thousands of data traces, and potential introduction of bias into the data analysis procedure, algorithms which identify the fingerprint unfolding length increments and classify the data correspondingly have been developed and implemented with success [69, 108–110].

1.7.2 Polymer elasticity models and contour length transformations

Single molecule force measurements generally gain access to a protein's extension x under a given force. The stochastic nature of domain unfolding or complex dissociation under force as well as the non-linear elastic behavior of the polymer backbone chain makes analysis in force-extension space difficult. The same unfolding event is observed over a range of different

positions in force-extension curves for multiple measurement cycles as shown in Fig. 1.4B i.

From a physicist's point of view, mechanical stretching of an unfolded protein domain is described by polymer elasticity models such as the worm-like chain (WLC) [111], the freely jointed chain (FJC) [112], or the freely rotating chain (FRC) model [113]. These models contain the free contour length L of the polymer, including surface tethers and unfolded protein backbone, as a parameter. The free contour length is simply the length of the polypeptide along the contour of the biopolymer chain, given a specific folding state (*e.g.*, Fig. 1.4A). Under a set of physically relevant constraints (L, x, $F > 0$, $x < L$), these elasticity models provide one-to-one mappings from force-extension space into force-contour length space. The models can be solved for the contour length parameter [71, 110], yielding an expression for the contour length as a function of force and extension $L(F, x)$. This function can be used to transform force-extension traces from constant speed or force clamp/ramp experiments into contour length space (Fig. 1.4B ii). The calculated contour length then can be binned (Fig. 1.4B iii), aligned, and subsequently averaged to precisely locate energy barriers (Fig. 1.4B iv) along a protein's unfolding pathway, and to classify data sets based on unfolding patterns. This idea was first proposed by Puchner *et al.* [110] and has been successfully applied in multiple AFM-SMFS studies [5, 69–71, 73, 114].

The WLC model accurately describes a protein's stretching response for forces up to approximately 200 pN. While many protein unfolding or dissociation events take place well within this force regime, some interactions like titin Ig domain unfolding [115], cohesin unfolding [116], dissociation of skeletal muscle titin-telethonin bonds [74] or dissociation of cellulosomal adhesion complexes [5, 72] exhibit much higher unfolding or rupture forces. To adequately describe the elastic response of polymers in such high force regimes, models beyond the standard WLC are required. To address this shortcoming, Hugel *et al.* [117] developed quantum mechanical corrections for polymer elasticity models to account for polypeptide backbone stretching at high forces. These corrections can be applied to obtain the contour length at zero force L_0 [110].

Livadaru *et al.* [113] proposed a more sophisticated model exhibiting three distinct regimes for a protein's stretching response as a function of the applied force F. It considers bonds of length b, connected by a fixed angle γ. The torsional angles are not restricted. The stretching behavior is then given by:

$$\frac{x}{L} = \begin{cases} \frac{Fa}{3k_\mathrm{B}T} & \text{for} \quad \frac{Fb}{k_\mathrm{B}T} < \frac{b}{p} \\ 1 - \left(\frac{4Fp}{k_\mathrm{B}T}\right)^{-\frac{1}{2}} & \text{for} \quad \frac{b}{p} < \frac{Fb}{k_\mathrm{B}T} < \frac{p}{b} \\ 1 - \left(\frac{2Fb}{k_\mathrm{B}T}\right)^{-1} & \text{for} \quad \frac{p}{b} < \frac{Fb}{k_\mathrm{B}T} \end{cases} \tag{1.1}$$

where $a = b\frac{1+\cos\gamma}{(1-\cos\gamma)\cos\frac{\gamma}{2}}$ is the Kuhn length, and $p = b\frac{\cos\frac{\gamma}{2}}{|\ln(\cos\gamma)|}$ is the effective persistence length. For AFM based SMFS, however, mainly the medium to high force regimes are relevant. The medium force regime of protein stretching, roughly between 10 pN − 125 pN, exhibits classical WLC stretching behavior, whereas the high force regime shows the behavior of a discrete chain, where the stretching response is independent of the persistence length. This model is most suitable for studying high force interactions, especially when combined with the aforementioned quantum mechanical corrections for backbone stretching. It has been used to analyze SMFS experiments described in Chapters 3 and 4.

1.7.3 Theoretical framework

The theoretical framework for single molecule force spectroscopy relies on seminal work of Bell [118] and was independently established by Evans and Ritchie [119] and Izrailev *et al.* [120] and published in the same issue of the *Biophysical Journal*. Protein unfolding or receptor ligand dissociation is modeled as diffusive crossing of a free energy barrier with height ΔG and distance to the barrier Δx, which is modulated by an external force. Assuming first order kinetics, and a monotonically increasing external force, the distribution of unfolding or rupture forces is given by:

$$p(F) = \frac{k(F)}{\dot{F}(F)} \exp\left(- \int_0^F \mathrm{d}F' \, \frac{k(F')}{\dot{F}(F')} \right) \tag{1.2}$$

where $k(F)$ is the force dependent off rate of the system, and $\dot{F}(F)$ is the force loading rate. In this picture the distribution of forces is solely defined by these two quantities, where $k(F)$ is defined by the underlying free energy landscape and $\dot{F}(F)$ is dictated by the experimental setup. Note that the force loading rate needs to be sufficiently slow such that intramolecular relaxation occurs much faster than the force induced barrier crossing for this fundamentally kinetic approach to remain valid [121, 122]. In the Bell [118] picture, the zero force off-rate

$$k_0 = \nu_0 \exp\left(\frac{\Delta G}{k_\mathrm{B} T} \right) \tag{1.3}$$

where ν_0 is the microscopic attempt frequency, is exponentially amplified under force and the distance to the free energy barrier remains constant:

$$k(F) = k_0 \exp\left(\frac{F \Delta x}{k_\mathrm{B} T} \right) \tag{1.4}$$

This constant barrier position model leads to an analytical expression for the distribution of unfolding or unbinding forces if a constant loading rate $\dot{F}(F) = \dot{F}$ is assumed:

$$p(F) = \frac{k_0}{\dot{F}} \exp\left[\frac{\Delta x}{k_\mathrm{B} T} F - \frac{k_0 k_\mathrm{B} T}{\Delta x \dot{F}} \left(\exp\left(\frac{\Delta x}{k_\mathrm{B} T} F \right) - 1 \right) \right] \tag{1.5}$$

In this model, the most probable rupture force can be found by formally setting $\frac{\mathrm{d}}{\mathrm{d}F} p(F) = 0$ and solving for F:

$$\langle F \rangle = \frac{k_\mathrm{B} T}{\Delta x} \ln\left(\frac{\Delta x \dot{F}}{k_0 k_\mathrm{B} T} \right) \tag{1.6}$$

Eq. (1.6) predicts a linear dependence of the most probable rupture force on the loading rate and is commonly referred to as the Bell-Evans formula. While this model is still widely used in the field of single molecule force spectroscopy, it has certain shortcomings which have been addressed in a number of theoretical works. Most prominently, the constant barrier position model assumes that the height of the energy barrier is reduced by $-Fx$, while its location Δx remains constant. Realistically, the distance to the transition state moves closer to the bound state as the energy landscape is tilted by the external force. Dudko *et al.* [123] have extended

this phenomenological picture by applying the Kramers theory [124] to different free energy landscapes to obtain a more sophisticated expression for the force dependent off-rate:

$$k(F) = k_0 \left(1 - \frac{\nu F \Delta x}{\Delta G}\right)^{\frac{1}{\nu}-1} \exp\left[\frac{\Delta G}{k_B T}\left(1 - \left(1 - \frac{\nu F \Delta x}{\Delta G}\right)^{\frac{1}{\nu}}\right)\right] \tag{1.7}$$

The model parameter ν defines the single-well free-energy surface model used in the Kramers approach ($\nu = \frac{2}{3}$ for linear-cubic and $\nu = \frac{1}{2}$ for cusp free-energy surface). For $\nu = 1$ and $\Delta G \to \infty$ independent of ν, the phenomenological model (Eqs. (1.4) and (1.5)) is recovered. Eq. (1.7) can in a next step be used in Eq. (1.2) to obtain the full distribution of forces for the constant loading rate case:

$$\begin{aligned} p(F) =& \frac{k(F)}{\dot{F}} \exp\left[\frac{k_0 k_B T}{\Delta x \dot{F}}\right] \\ & \times \exp\left[-\left(\frac{k(F) k_B T}{\Delta x \dot{F}}\right)\left(1 - \frac{\nu F \Delta x}{\Delta G}\right)^{1-\frac{1}{\nu}}\right] \end{aligned} \tag{1.8}$$

This result then yields an asymptotic expression for the mean rupture force $\langle F \rangle = \int \mathrm{d}F\, F p(F)$ as a function of the loading rate [123, 125]:

$$\langle F \rangle = \frac{\Delta G}{\nu \Delta x}\left(1 - \left(\frac{k_B T}{\Delta G} \ln\left(\frac{k_0 k_B T}{\Delta x \dot{F}} \exp\left[\frac{\Delta G}{k_B T} + \gamma\right]\right)\right)^{\nu}\right) \tag{1.9}$$

where $\gamma = 0.577\ldots$ is the Euler-Mascheroni constant. This non-linear dependence of the mean rupture force on the logarithm of the loading rate has been observed in a number of experimental studies (for a list of such data sets see Friddle *et al.* [122]). The Dudko-Hummer-Szabo model has for example been used to analyze high speed AFM data bridging the gap between experiments and molecular dynamics simulations [95] and in Chapter 4 to analyze joint experimental and computational studies. It has to be noted that the assumption of a constant loading rate $\dot{F}$ common to the aforementioned theoretical concepts is experimentally only realized in a force ramp protocol [126]. However, velocity clamp, *i.e.* constant speed measurements, is the experimentally more common pulling protocol. In constant speed pulling, the force loading rate usually becomes a function of the applied force due to the presence of molecular linkers with non-linear elasticity. The force dependent loading rate $\dot{F}(F)$ can be calculated by applying a simple force balance:

$$\frac{v}{\dot{F}} = \frac{1}{k} + \frac{\partial x}{\partial F} \tag{1.10}$$

where v, k, and $x(F)$ are the pulling speed, the spring constant of the harmonic pulling device, and the force dependent extension of the linker, respectively. Eq. (1.10) can be used to obtain an expression for $\dot{F}(F)$ by accounting for molecular linkers using an appropriate polymer elasticity model. Dudko *et al.* [121] used Eq. (1.10) to derive the force dependent loading rate for a worm-like chain linker (Eq. (4) in ref. [121]). For forces beyond the WLC regime, Eq. (1.1) can be used to obtain an analogue expression given the high force stretching response proposed by Livadaru *et al.* [113].

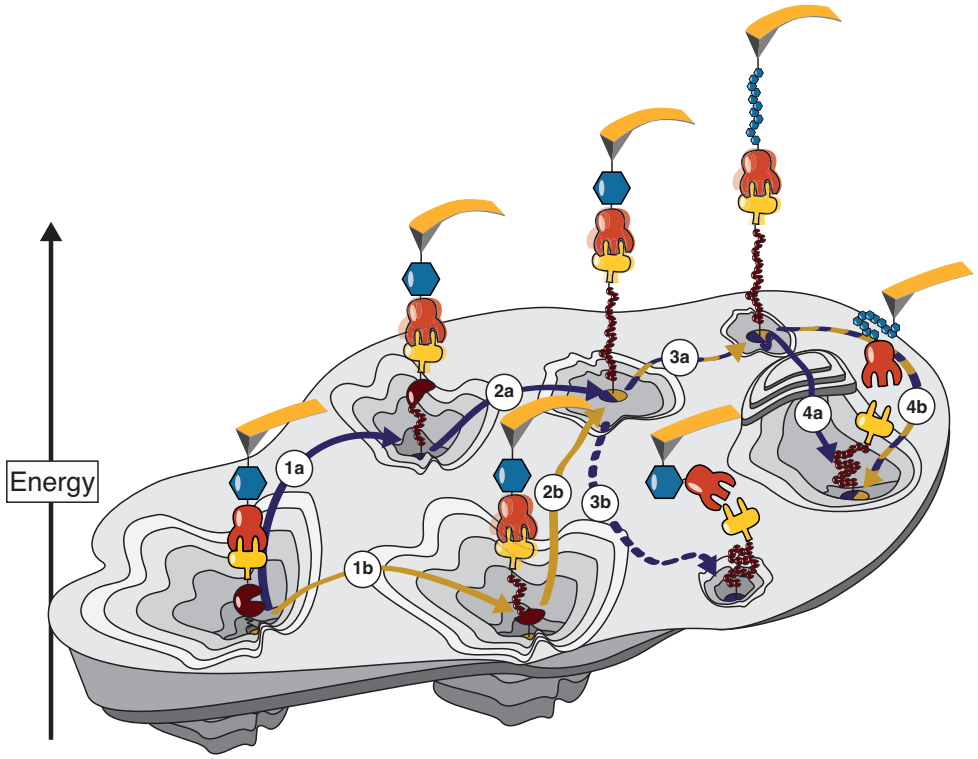


Figure 1.5: Schematic depiction of an (un)folding energy landscape. The bound state of a protein receptor-ligand complex can be thought of as a Brownian particle confined to a complex multidimensional energy landscape. At equilibrium, the system can escape the bound state driven by thermal fluctuations. This escape can occur along any pathway on the energy landscape. When measuring the thermal off-rate with bulk assays such as surface plasmon resonance biosensors, a weighted average of all thermally accessible pathways is obtained. In a single-molecule pulling experiment, however, a small subset of pathways is selected, which is defined by the projection of the energy landscape onto the pulling coordinate as illustrated by paths 1-3. Caution is required when comparing data obtained from single molecule techniques with bulk data. In cases where SMFS probes a steep pathway with a high free energy barrier, the fitted zero-force off rate may greatly differ from values obtained by bulk techniques. Path 4 illustrates the thermal escape (4b) versus the forced pathway across an additional energy barrier (4a) by the AFM cantilever.

Although in both bulk and single molecule force measurements at common loading rates, the unbinding process is thermally driven, caution is required when comparing their results. At unbiased equilibrium, all thermally accessible pathways from the bound state are sampled and the off rate is consequently measured as a weighted average. Single molecule force measurements, however, select only a small subset of these pathways due to the defined pulling geometry, as illustrated by paths 1-3 in Fig. 1.5. In cases where the energy landscape is highly asymmetric and the pulling experiment probes a steep pathway, the off rates obtained from single molecule *vs.* bulk measurements might differ greatly (see Fig. 1.5, paths 4a vs 4b). This behavior is reflected in common place biochemical affinities *vs.* exceptional mechanical strength along these unbinding pathways and is further discussed in Chapter 3.

1.8 Outlook

In the future, there remain several technical challenges that need to be addressed. One of the limitations of AFM is that it covers a relatively high force range, yet there exist a multitude of biological interactions in the low-force regime that are of interest. Further technical advances in instrument design, cantilever fabrication, and feedback control might further improve force resolution and thereby enable such experiments. A second area for improvement involves sample throughput and parallel screening. With the development of more elaborate, sophisticated and well defined surface immobilization strategies and protein handles, significant gains in throughput can be envisioned. Innovations of the chemistry in combination with efficient data analysis protocols and state of the art instrumentation may pave the way towards in depth study of complex, multi-domain protein systems.

These advances in experimental design and throughput would greatly benefit from refined theoretical frameworks that account for parameters such as cantilever stiffness and dampening whilst maintaining analytical tractability. Consequently, with improved methodology we anticipate the community will be able to address an even wider range of questions about mechanical adaptations of proteins and protein complexes in the future.

Chapter 2

Molecular Dynamics Simulations

2.1 Introduction

Molecular dynamics simulations provide a means to study the dynamics of molecular systems *in silico*. The dynamic evolution of a structure, *e.g.* from x-ray crystallography, is obtained by solving Newton's equation of motion for all atoms of the system

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial}{\partial \mathbf{r}_i} U\left(\mathbf{r}_1, \mathbf{r}_2 \ldots \mathbf{r}_N\right) = -\frac{\partial}{\partial \mathbf{r}_i} U\left(\mathbf{R}\right) \tag{2.1}$$

where m_i and $\mathbf{r}_i$ are the mass and position of atom i, respectively, and $U\left(\mathbf{R}\right)$ is the potential energy of the system, which is dependent on the positions of all atoms in the system. In practice, Eq. (2.1) is numerically integrated using Verlet's algorithm [127]. This method considers a Taylor expansion of an atomic position $\mathbf{r}_i\left(t\right)$ with a time step δt forward and backward in time:

$$\mathbf{r}_i\left(t+\delta t\right) \approx \mathbf{r}_i\left(t\right) + \dot{\mathbf{r}}_i\left(t\right)\delta t + \frac{1}{2}\ddot{\mathbf{r}}_i\left(t\right)\delta t^2 \tag{2.2}$$

$$\mathbf{r}_i\left(t-\delta t\right) \approx \mathbf{r}_i\left(t\right) - \dot{\mathbf{r}}_i\left(t\right)\delta t + \frac{1}{2}\ddot{\mathbf{r}}_i\left(t\right)\delta t^2 \tag{2.3}$$

Addition of Eqs. (2.2) and (2.3) and reordering yields a recursive formula for the time evolution of the system:

$$\mathbf{r}_i\left(t+\delta t\right) \approx 2\mathbf{r}_i\left(t\right) - \mathbf{r}_i\left(t-\delta t\right) + \ddot{\mathbf{r}}_i \delta t^2 \tag{2.4}$$

Initially, every atom is assigned a random initial velocity $\dot{\mathbf{r}}_i$ drawn from a Maxwell-Boltzmann distribution:

$$f\left(v\right) = 4\pi v^2 \sqrt{\left(\frac{m}{2\pi k_\mathrm{B} T}\right)^3} \exp\left(-\frac{m v^2}{2 k_\mathrm{B} T}\right) \tag{2.5}$$

and Eq. (2.2) is used to perform the first integration step.

Eq. (2.1) states that the entire dynamics of the system is described by the energy function $U\left(\mathbf{R}\right)$, which is captured in so called force fields in molecular dynamics calculations. These force fields, including the most commonly used CHARMM [128] field, combine a set of heuristic terms

describing bonded interactions and terms depicting non-bonded interactions, *i.e.* van-der-Waals and electrostatic interactions:

$$\begin{aligned} U(\mathbf{R}) &= \sum_{\text{bonds}} k_{\text{b}} (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\ &+ \sum_{\text{dihedrals}} k_\phi (1 + \cos(n\phi + \delta)) + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 \\ &+ \sum_{\text{Urey-Bradley}} k_{\text{UB}} (u - u_0)^2 \\ &+ \sum_{\text{non-bonded}} \epsilon \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_M r_{ij}} \end{aligned} \tag{2.6}$$

The first term in Eq. (2.6) describes oscillations of two bonded atoms about an equilibrium bond length b_0 with force constant k_{b}. The second term describes bending oscillations of three bonded atoms about their equilibrium bond angle θ_0 with an angular force constant k_θ. The third term accounts for dihedrals (or torsional rotations) with dihedral angle ϕ and force constant k_ϕ, respectively, and multiplicity n and phase shift δ. The forth term describes out of plane bending with an out of plane angle $\omega - \omega_0$ with a corresponding force constant k_ω. The Urey-Bradley term introduces a virtual bond between the first and third atom that are bonded to a middle atom (think between the hydrogens in a water molecule). It is used on a case by case basis, like the impropers term, to optimize force field predictions to data from vibrational spectra [128]. Non-bonded interactions, namely van-der-Waals and electrostatic interactions, are represented by a 12-6 Lennard-Jones and coulomb potential, respectively.

Molecular dynamics simulations have been referred to as a computational microscope [129] since they provide insight into dynamic processes with atomistic detail, especially when combined with structural information from experiments such as cryo electron microscopy (EM) data. With the rise of supercomputers, impressive results have been achieved in recent years. In 2006, Jayachandran *et al.* [130] reported the first all atom simulation of a complete folding trajectory of the villin headpiece, a 36-residue α-helical protein. Only a few years later, combined cryo EM and molecular dynamics studies were able to resolve a full atomistic model of the mature human immunodeficiency virus-1 (HIV-1) capsid [131] and show how the antibiotic erythromycin allosterically introduces translation arrest at bacterial ribosomes [132]. A comprehensive overview of molecular dynamics studies of macromolecular systems was recently given by Perilla *et al.* [133].

2.2 Steered molecular dynamics

Steered molecular dynamics expand the computational toolbox by adding external forces to molecular dynamics simulations. While in SMD the force protocol applied to a system is in principle arbitrary, the most commonly used methods are constant velocity (CV) and constant force (CF) pulling, since they are the most frequently used methods in experimental single molecule force spectroscopy. Among the first reports of SMD were simulated unbinding studies of biotin streptavidin [134] and biotin avidin [120]. Especially when combined with pulling data from experiments, steered molecular dynamics can reveal unbinding or unfolding mechanisms

with unparalleled detail. In Chapter 3, single molecule force spectroscopy and steered molecular dynamics were combined to investigate the exceptional mechanostability of a cellulosomal receptor ligand complex from the ruminal bacterium *Ruminococcus flavefaciens*.

2.3 Force propagation analysis

In Chapter 4 the type III cohesin dockerin complex studied in Chapter 3 was used as model system to investigate the directional dependence of receptor ligand unbinding and to analyze the paths along which an external force propagates through a molecular system. To this end, a new analysis protocol for molecular dynamics trajectories was established that combines thermodynamic fluctuation theory and tools from network based correlation analysis. A basic result from thermodynamic fluctuation theory relates the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i [135, 136]:

$$\left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle = k_\mathrm{B} T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{2.7}$$

where $\Delta\mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$, and $\mathbf{r}_i$ is the position of atom i. Eq. (2.7) supports an intuitive result: Two atoms i and j will move in a correlated fashion due to an external force, if the interaction potential between them is strong. Under such an external force, soft degrees of freedom are expected to be "stretched out" and the dynamics of the system are determined by stiff paths through the molecule, *i.e.* force will propagate through a molecular structure along these stiff paths. In chapter 4, a simple toy model of atoms of identical mass connected by harmonic springs of different stiffnesses was used to validate this picture (Fig. 2.1A). The NAMD [137] SMD [120] CV software package was used to pull on one of the green atoms (Fig. 2.1B, arrow) while the second green atom (Fig. 2.1B, anchor) was fixed. The resulting molecular dynamics trajectory was used to calculate a correlation matrix C_{ij} based on the Pearson correlation coefficient. The elements from this matrix were then used to weight the connections between the atoms in the toy model (Fig. 2.1). The Pearson correlation coefficient C_{ij} is given by:

$$C_{ij} = \frac{\left\langle \Delta\mathbf{r}_i(t) \cdot \Delta\mathbf{r}_j(t) \right\rangle}{\left(\left\langle \Delta\mathbf{r}_i(t)^2 \right\rangle \left\langle \Delta\mathbf{r}_j(t)^2 \right\rangle \right)^{\frac{1}{2}}} \tag{2.8}$$

Combining Eqs. (2.7) and (2.8) yields:

$$C_{ij} = k_\mathrm{B} T \,\mathrm{Tr}\, \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\left\langle \Delta\mathbf{r}_i^2(t) \right\rangle \left\langle \Delta\mathbf{r}_j^2(t) \right\rangle \right)^{-\frac{1}{2}} \tag{2.9}$$

Assuming an arbitrary potential $U_i(\mathbf{r})$ of atom i, a Taylor expansion around the minimum (set to be at zero w.l.o.g.) yields:

$$U_i(\mathbf{r}) = 0 + \underbrace{\mathbf{r}_i^T \nabla U(0)}_{=0} + \frac{1}{2} \mathbf{r}_i^T \underline{H}(0)\, \mathbf{r}_i + \ldots \tag{2.10}$$

where $\underline{H}(0)$ is the Hessian matrix evaluated at the potential minimum. Assuming Schwarz' theorem holds for $U_i(\mathbf{r})$, $\underline{H}(0)$ is a symmetric matrix and therefore has real eigenvalues and

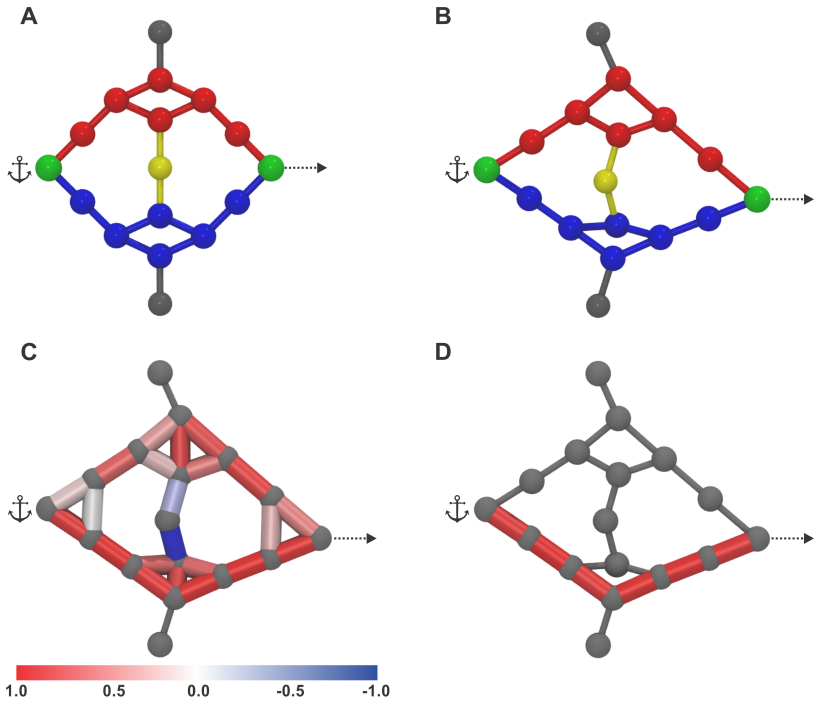


Figure 2.1: Force Propagation Analysis - toy model. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red: k, blue: $5k$, yellow: $7.5k$, black: $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs where introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red. Figure and caption adapted from Schoeler *et al.* [72] with permission from the American Chemical Society.

orthonormal eigenvectors. Hence, a change to the eigenbasis of $\underline{H}(0)$ is a rotation of the coordinate system. In this new basis the Hessian is diagonal:

$$\underline{H}(0) \rightarrow \underline{H}'(0) = \begin{pmatrix} k_{x'} & 0 & 0 \\ 0 & k_{y'} & 0 \\ 0 & 0 & k_{z'} \end{pmatrix} \tag{2.11}$$

This yields a simple expression for the second order term in Eq. (2.10):

$$U_i(\mathbf{r}') = \frac{1}{2}\mathbf{r}'^T \underline{H}'(0)\,\mathbf{r}' = \frac{1}{2}\left(k_{x'}x'^2 + k_{y'}y'^2 + k_{z'}z'^2\right) \tag{2.12}$$

Now we inspect the right hand side of equation (2.9):

$$\langle \Delta\mathbf{r}_i^2(t)\rangle = \langle \mathbf{r}_i^2(t) - 2\mathbf{r}_i(t)\langle \mathbf{r}_i(t)\rangle + \langle \mathbf{r}_i(t)\rangle^2\rangle \tag{2.13}$$

In the harmonic approximation of the potential of atom i, $\langle \mathbf{r}_i(t)\rangle = 0$, and therefore $\langle \Delta\mathbf{r}_i(t)^2\rangle = \langle \mathbf{r}_i^2(t)\rangle$. In the basis of $\underline{H}'(0)$ this becomes:

$$\langle \mathbf{r}_i'^2(t)\rangle = \langle x_i'(t)^2 + y_i'(t)^2 + z_i'(t)^2\rangle = \langle x_i'(t)^2\rangle + \langle y_i'(t)^2\rangle + \langle z_i'(t)^2\rangle \tag{2.14}$$

Applying the equipartition theorem to this result yields:

$$\langle x_i'(t)^2\rangle = \frac{k_\mathrm{B}T}{k'_{x_i}} \tag{2.15}$$

And therefore:

$$\langle \Delta\mathbf{r}_i'(t)^2\rangle = k_\mathrm{B}T\left(\frac{1}{k'_{x_i}} + \frac{1}{k'_{y_i}} + \frac{1}{k'_{z_i}}\right) = \frac{k_\mathrm{B}T}{k'_{i,\mathrm{eff}}} \tag{2.16}$$

Plugging this result into Eq. (2.9), one finds:

$$C_{ij} = k_\mathrm{B}T\,\mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\left(\frac{k_\mathrm{B}T}{k'_{i,\mathrm{eff}}}\right)^{-\frac{1}{2}}\left(\langle\Delta\mathbf{r}_j(t)^2\rangle\right)^{-\frac{1}{2}} \tag{2.17}$$

Repeating the above steps for atom j yields the final result:

$$C_{ij} = k_\mathrm{B}T\,\mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\left(\frac{k_\mathrm{B}T}{k'_{i,\mathrm{eff}}}\right)^{-\frac{1}{2}}\left(\frac{k_\mathrm{B}T}{k'_{j,\mathrm{eff}}}\right)^{-\frac{1}{2}} \tag{2.18}$$

$$= \mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\sqrt{k'_{i,\mathrm{eff}}\cdot k'_{j,\mathrm{eff}}} \tag{2.19}$$

These calculations illustrate how the Pearson correlation coefficient is indeed a suitable measure for the identification of stiff, force bearing pathways through a biomolecular structure. For the toy model, the path of highest correlation was identified using dynamical network analysis [138], and expectedly turned out to be the path connected by the stiffest springs (Fig. 2.1). The same methodology was then applied to the cohesin dockerin model system to study force propagation pathways and their implications for mechanostability in different pulling geometries. The results of this study are summarized in Chapter 4. In brief, it was found that this system locally directs mechanical force across the binding interface at favorable angles in cases where exceptional mechanostability is observed, as shown in a detailed and schematic representation in Fig. 2.2. Whether this is a mechanism unique to cellulosomal receptor ligand complexes or a more general mechanism for mechanostability will remain a question for further research.

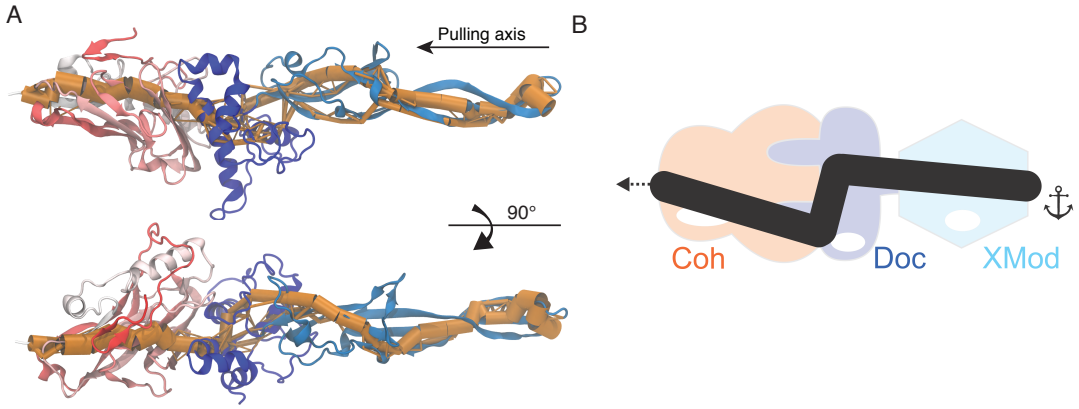


Figure 2.2: Force Propagation Analysis - model cohesin dockerin system. (A) Network paths for the loaded cohesin dockerin system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (B) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large normal components to the unbinding axis.
Figure and caption adapted from Schoeler *et al.* [72] with permission from the American Chemical Society.

Part II

Results

Chapter 3

Mechanics of an ultrastable cellulosome adhesion complex

Summary

Challenging environments have guided nature in the development of ultrastable protein complexes. Specialized bacteria produce discrete multi-component protein networks called cellulosomes to effectively digest lignocellulosic biomass. While network assembly is enabled by protein interactions with commonplace affinities, we show that certain cellulosomal receptor-ligand interactions exhibit extreme resistance to applied force. Here, we characterized the ligand-receptor complex responsible for substrate anchoring in the *Ruminococcus flavefaciens* cellulosome using single-molecule force spectroscopy and steered molecular dynamics simulations. The complex withstands forces of 600-750 pN, situating it among the strongest bimolecular interactions reported, equivalent to half the mechanical strength of a covalent bond. Our findings demonstrate force activation and inter-domain stabilization of the complex, and suggest that certain network components serve as mechanical effectors for maintaining network integrity. This detailed understanding of cellulosomal network components can help in the development of biocatalysts for production of fuels and chemicals from renewable plant-derived biomass.

3.1 Introduction

Cellulosomes are protein networks designed by nature to degrade lignocellulosic biomass [139]. These networks comprise intricate assemblies of conserved subunits including catalytic domains, scaffold proteins, carbohydrate binding modules (CBMs), cohesins (Cohs), dockerins (Docs), and X-modules (XMods) of unknown function. Coh:Doc pairs form complexes with high affinity and specificity [140], and provide connectivity to a myriad of cellulosomal networks with varying Coh-Doc network topology [141–143]. The most intricate cellulosome known

to date is produced by *Ruminococcus flavefaciens* [144, 145] and contains several primary and secondary scaffolds along with over 220 dockerin-bearing protein subunits [146].

The importance of cellulolytic enzymes for the production of renewable fuels and chemicals from biomass has highlighted an urgent need for improved fundamental understanding of how cellulosomal networks achieve their impressive catalytic activity [147]. Two of the mechanisms known to increase the catalytic activity of cellulosomes are proximity and targeting effects [148]. Proximity refers to the high local concentration of enzymes afforded by incorporation into nanoscale networks, while targeting refers to specific binding of cellulosomes to substrates. Protein scaffolds and CBM domains are both critical in this context as they mediate interactions between comparatively large bacterial cells and cellulose particles. Since many cellulosomal habitats (*e.g.*, cow rumen) exhibit strong flow gradients, shear forces will accordingly stress bridging scaffold components mechanically *in vivo*. Protein modules located at stressed positions within these networks should therefore be preselected for high mechanostability. However, thus far very few studies on the mechanics of carbohydrate-active proteins or cellulosomal network components have been reported[149].

In the present study we sought to identify cellulosomal network junctions with maximal mechanical stability. We chose an XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose in the rumen. The complex links the *Ruminococcus flavefaciens* cell wall to the cellulose substrate via two CBM domains located at the N-terminus of the CttA scaffold, as shown in Fig. 3.1a. The crystal structure of the complex solved by X-ray crystallography [4] is shown in Fig. 3.1b. XMod-Doc tandem dyads such as this one are a common feature in cellulosomal networks. Bulk biochemical assays on XMod-Docs have demonstrated that XMods improve Doc solubility and increase biochemical affinity of Doc:Coh complex formation [150]. Crystallographic studies conducted on XMod-Doc:Coh complexes have revealed direct contacts between XMods and their adjacent dockerins [4, 151]. Additionally, many XMods (*e.g.*, PDB 2B59, 1EHX, 3PDD) have high β-strand content and fold with N- and C-termini at opposite ends of the molecule, suggestive of robust mechanical clamp motifs at work [152, 153]. These observations are all suggestive of a mechanical role for XMods.

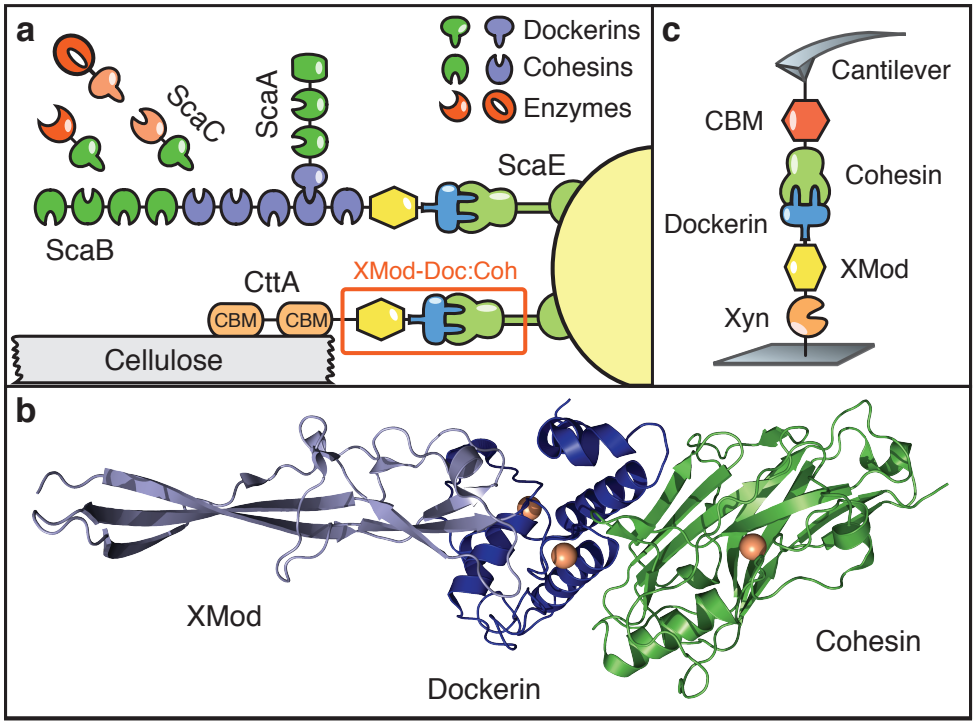


Figure 3.1: System overview. (a) Schematic of selected components of the *R. flavefaciens* cellulosome. The investigated XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose is highlighted in orange. (b) Crystal structure of the XMod-Doc:Coh complex. Ca^{2+} ions are shown as orange spheres. (c) Depiction of experimental pulling configuration I, with CBM-Coh attached to the cantilever tip and Xyn-XMod-Doc attached to the glass surface.

3.2 Results

We performed single molecule force spectroscopy (SMFS) experiments with an atomic force miscroscope (AFM) to probe the mechanical dissociation of XMod-Doc:Coh. Xylanase (Xyn) and CBM fusion domains on the XMod-Doc and Coh modules, respectively, provided identifiable unfolding patterns permitting screening of large datasets of force-distance curves [71, 154]. Engineered cysteines and/or peptide tags on the CBM and Xyn marker domains were used to covalently immobilize the binding partners in a site-specific manner to an AFM cantilever or cover glass via poly(ethylene glycol) (PEG) linkers. The pulling configuration with Coh-CBM immobilized on the cantilever is referred to as configuration I, as shown in Fig 3.1c. The reverse configuration with Coh-CBM on the cover glass is referred to as configuration II. In a typical experimental run we collected about 50,000 force extension traces from a single cantilever. We note that the molecules immobilized on the cantilever and glass surfaces were stable over thousands of pulling cycles.

We sorted the data by first searching for contour length increments that matched our specific xylanase and CBM fingerprint domains. After identifying these specific traces (Fig. 3.2a), we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history. To assign protein subdomains to the observed unfolding patterns, we transformed the data into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching (QM-FRC, Supplementary Note 2, Supplementary Fig. 3.5) [113, 117]. The fit parameter-free QM-FRC model describes protein stretching at forces greater than 200 pN more accurately than the commonly used worm-like chain (WLC) model [110, 113]. The resulting contour length histogram is shown in Fig. 3.2b. Peak-to-peak distances in the histogram represent contour length increments of unfolded protein domains. Assuming a length per stretched amino acid of 0.365 nm and accounting for the folded length of each subdomain, we compared the observed increments to the polypeptide lengths of individual subdomains of the Xyn-XMod-Doc and Coh-CBM fusion proteins. Details on contour length estimates and domain assignments are shown in Supplementary Table 3.1.

Unfolding patterns in configuration I showed PEG stretching followed by a three-peaked Xyn fingerprint (Fig. 3.2a, top trace, green), which added 90 nm of contour length to the system. Xyn unfolding was followed by CBM unfolding at $\sim$ 150 pN with 55 nm of contour length added. Finally, the XMod-Doc:Coh complex dissociated at an ultra-high rupture force of $\sim$ 600 pN. The loading rate dependence of the final rupture event for curves of subtype 1 is plotted in Fig. 3.2c (blue). The measured complex rupture force distributions are shown in Fig. 3.6.

Less frequently (35-40% of traces) we observed a two-step dissociation process wherein the XMod unfolded prior to Doc:Coh rupture as shown in Fig. 3.2a, (middle trace, orange). In these cases, the final dissociation exhibited a much lower rupture force ($\sim$ 300 pN) than the preceding XMod unfolding peak, indicating the strengthening effect of XMod was lost, and XMod was no longer able to protect the complex from dissociation at high force. The loading rate dependency of Doc:Coh rupture occurring immediately following XMod unfolding is shown in Fig. 3.2c (gray).

In configuration II (Fig. 3.2a, bottom trace), with the Xyn-XMod-Doc attached to the cantilever, the xylanase fingerprint was lost after the first few force extension traces acquired in the dataset. This indicated the Xyn domain did not refold within the timescale of the experiment once unfolded, consistent with prior work[71, 154]. CBM and XMod unfolding events were

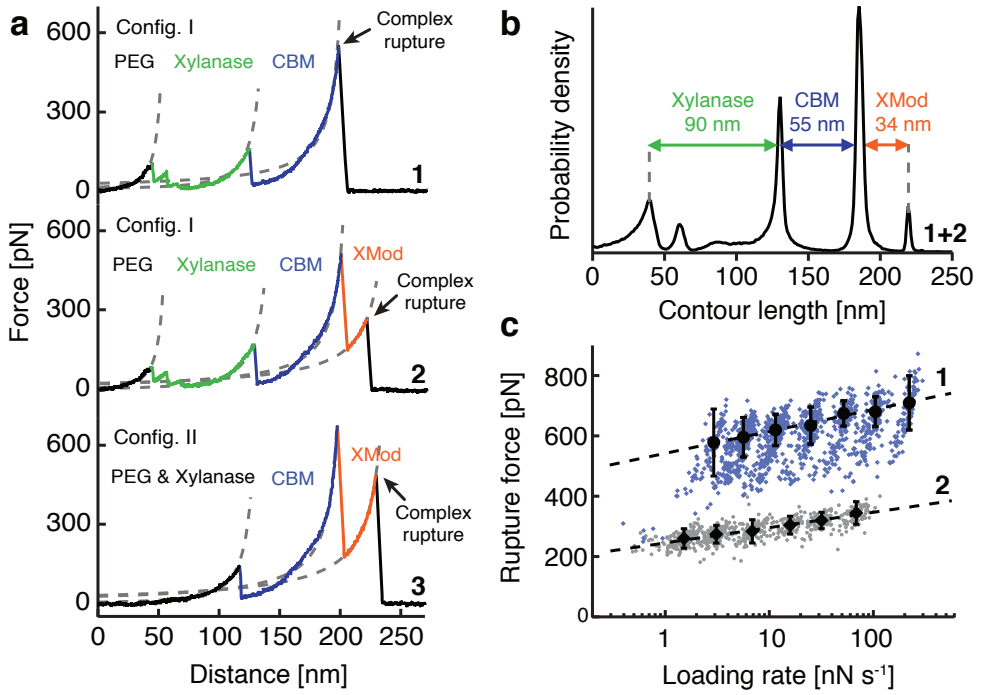


Figure 3.2: Experimental SMFS unfolding traces. (a) Unfolding fingerprints from pulling configuration I (curves 1 & 2) and configuration II (curve 3). The QM-FRC model (dashed lines) was used to estimate the contour lengths of the unfolded modules. (b) Contour length histogram obtained from 127 force extension traces (Config. I). The peak-to-peak increments correspond to Xyn, CBM and XMod amino acid sequence lengths. (c) Dynamic force spectra for the final Doc:Coh complex rupture peaks obtained from 2122 force-extension traces. The blue points show Doc:Coh ruptures that occurred with an intact XMod, while gray points show ruptures immediately following XMod unfolding. Black circles and diamonds represent the most probable rupture force/loading rate obtained by Gaussian fitting at each pulling speed. Error bars are ±1 standard deviation. Dashed lines are least squares fits to the Bell-Evans model.

observed repeatedly throughout the series of acquired force traces in both configurations I and II, indicating these domains were able to refold while attached to the cantilever over the course of the experiment.

We employed the Bell-Evans model[155] (Supplementary Note 3) to analyze the final rupture of the complex through the effective distance to the transition state (Δx) and the natural off-rate (k_{off}). The fits to the model yielded values of $\Delta x = 0.13\,\text{nm}$ and $k_{\text{off}} = 7.3 \times 10^{-7}\,\text{s}^{-1}$ for an intact Xmod, and $\Delta x = 0.19\,\text{nm}$ and $k_{\text{off}} = 4.7 \times 10^{-4}\,\text{s}^{-1}$ for the "shielded" rupture following XMod unfolding (Fig. 3.2c). These values indicate that the distance to the transition state is increased following XMod unfolding, reflecting an overall softening of the binding interface. Distances to the transition state observed for other receptor-ligand pairs are typically on the order of $\sim 0.7\,\text{nm}$ [154]. The extremely short Δx of 0.13 nm observed here suggests that mechanical unbinding for this complex is highly coordinated. We further analyzed the unfolding of XMod in the Bell-Evans picture and found values of $\Delta x = 0.15\,\text{nm}$ and $k_{\text{off}} = 2.6 \times 10^{-6}\,\text{s}^{-1}$. The loading rate dependence for this unfolding event is shown in Fig. 3.7.

The exceptionally high rupture forces measured experimentally (Fig. 3.2) are hugely disproportionate to the XMod-Doc:Coh biochemical affinity, which at $K_{\text{D}} \sim 20\,\text{nM}$ [4] is comparable to typical antibody-antigen interactions. Antibody-antigen interactions, however, will rupture at only $\sim 60\,\text{pN}$ at similar loading rates [156], while bimolecular complexes found in muscle exposed to mechanical loading in vivo will rupture at $\sim 140\,\text{pN}$ [157]. Trimeric titin-telethonin complexes also found in muscle exhibit unfolding forces around 700 pN[74] while Ig domains from cardiac titin will unfold at $\sim 200\,\text{pN}$ [158]. The XMod-Doc:Coh ruptures reported here fell in a range from $600\,\text{pN} - 750\,\text{pN}$ at loading rates ranging from $10\,\text{nN}\,\text{s}^{-1} - 100\,\text{nN}\,\text{s}^{-1}$. At around half the rupture force of a covalent gold-thiol bond [159], these bimolecular protein rupture forces are, to the best of our knowledge, among the highest of their kind ever reported. The covalent bonds in this system are primarily peptide bonds in the proteins and C-C and C-O bonds in the PEG linkers. These are significantly more mechanically stable than the quoted gold-thiol bond rupture force ($\sim 1.2\,\text{nN}$) [160] and fall in a rupture force range $> 2.5\,\text{nN}$ at similar loading rates. Therefore, breakage of covalent linkages under our experimental conditions is highly unlikely. We note that the high mechanostability observed here is not the result of fusing the proteins to the CBM or Xyl domains. The covalent linkages and pulling geometry are consistent with the WT complex and its dissociation pathway. *In vivo*, the Coh is anchored to the peptidoglycan cell wall through its C-terminal sortase motif. The XMod-Doc is attached to the cellulose substrate through two N-terminal CBM domains. By pulling the XMod-Doc from through an N-terminal Xyl fusion domain, and the Coh through a C-terminal CBM, we established an experimental pulling geometry that matches loading of the complex *in vivo*. This pulling geometry was also used in all simulations. The discontinuity between its commonplace biochemical affinity and remarkable resistance to applied force illustrates how this complex is primed for mechanical stability, and highlights differences in the unbinding pathway between dissociation at equilibrium and dissociation induced mechanically along a defined pulling coordinate.

To elucidate the molecular mechanisms at play that enable this extreme mechanostability, we carried out all-atom SMD simulations. The Xyn and CBM domains were not modeled in order to keep the simulated system small and reduce the usage of computational resources. This approximation was reasonable since we have no indication that these domains significantly affect the XMod-Doc:Coh binding strength [161]. After equilibrating the crystal structure [4],

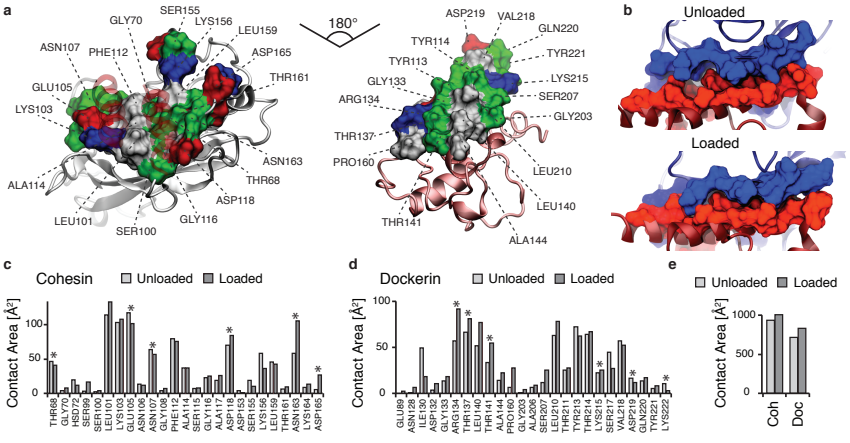


Figure 3.3: Analysis of binding interface and catch bond mechanism from SMD. (a) Surface plots for the main interacting residues of Coh (left) and Doc (right). Hydrophobic residues are shown in gray, polar residues in green, and negative and positive residues in red and blue, respectively. Both Coh and Doc exhibit a hydrophobic patch in the center of the binding surface that is surrounded by polar and charged residues. (b) Rearrangement of binding residues of Coh (blue) and Doc (red) under force. Following mechanical loading, an interdigitated complex is formed that resembles teeth of a zipper. (c,d) Surface contact area of interacting residues of Coh (c) and Doc (d) in the absence and presence of force. Residues forming prevalent hydrogen bonds are indicated with stars. (e) Total contact surface area of Coh and Doc in unloaded and loaded conformations.

the N-terminus of XMod-Doc was harmonically restrained while the C-terminus of Coh was pulled away at constant speed. The force applied to the harmonic pulling spring was stored at each time-step. We tested pulling speeds of 0.25, 0.625 and 1.25 A ns^{-1}, and note that the slowest simulated pulling speed was ~4,000 times faster than our fastest experimental pulling speed of 6.4 µm s^{-1}. This difference is considered not to affect the force profile, but it is known to account for the scale difference in force measured by SMD and AFM [134, 162].

SMD results showed the force increased with distance until the complex ruptured for all simulations. At the slowest pulling speed of 0.25 A ns^{-1} the rupture occurred at a peak force of ~ 900 pN, as shown in Supplementary Fig. 3.8 and Supplemental Movie 1. We analyzed the progression and prevalence of hydrogen bonded contacts between the XMod-Doc and Coh domains to identify key residues in contact throughout the entire rupture process and particularly immediately prior to rupture. These residues are presented in Fig. 3.3a, c, d and Supplementary Figs. 3.9 and 3.10. The simulation results clearly reproduced key hydrogen bonding contacts previously identified[4] as important for Doc:Coh recognition (Supplementary Fig. 3.9).

The main interacting residues are shown in Fig. 3.3a and b. Both Coh and Doc exhibit a binding interface consisting of a hydrophobic center (gray) surrounded by a ring of polar (green) and charged residues (blue, positive; red, negative). This residue pattern suggests the hydrophilic side chains protect the interior hydrophobic core from attack by water molecules, compensating for the flat binding interface that lacks a deep pocket. The geometry suggests a penalty to unbinding that stabilizes the bound state. Further, we analyzed the contact surface areas of interacting residues (Fig. 3.3b-e). The total contact area was found to increase due to rearrangement of the interacting residues when the complex is mechanically stressed, as shown in Fig. 3.3e and Movie S2. Doc residues in the simulated binding interface clamped down on Coh residues upon mechanical loading, resulting in increased stability and decreased

accessibility of water into the hydrophobic core of the bound complex (Fig. 3.3b). These results suggest that a catch bond mechanism is responsible for the remarkable stability [163] under force and provide a molecular mechanism which the XMod-Doc:Coh complex uses to summon mechanical strength when needed, while still allowing relatively fast assembly and disassembly of the complex at equilibrium. The residues that increase most in contact area (Fig. 3.33c, d) present promising candidates for future mutagenesis studies.

Among the 223 Doc sequences from *Ruminococcus flavefaciens*, 6 subfamilies have been explicitly identified using bioinformatics approaches [146]. The XMod-Doc investigated here belongs to the 40-member Doc family 4a. A conserved feature of these Doc modules is the presence of 3 sequence inserts that interrupt the conserved duplicated F-hand motif Doc structure. In our system, these Doc sequence inserts make direct contacts with XMod in the crystallized complex (Fig. 3.1), and suggest an interaction between XMod and Doc that could potentially propagate to the Doc:Coh binding interface. To test this, an independent simulation was performed to unfold XMod (Fig. 3.4). The harmonic restraint was moved to the C-terminus of XMod so that force was applied from the N- to C-terminus of XMod only, while leaving Doc and Coh unrestrained. The results (Fig. 3.4b) showed XMod unfolded at forces slightly higher than but similar to the Xmod-Doc:Coh complex rupture force determined from the standard simulation at the same pulling speed. This suggested XMod unfolding prior to Doc:Coh rupture was not probable, but could be observed on occasion due to the stochastic nature of domain unfolding. This was consistent with experiments where XMod unfolding was observed in $\sim 35\,\% - 40\,\%$ of traces. Furthermore, analysis of the H-bonding between Doc and XMod (Fig. 3.4d, red) indicated loss of contact as Xmod unfolded, dominated by contact loss between the 3 Doc insert sequences and XMod. Interestingly, XMod unfolding clearly led to a decrease in H-bonding between Doc and Coh at a later stage (~ 200 ns) well after XMod had lost most of its contact with Doc, even though no force was being applied across the Doc:Coh binding interface. This provided evidence for direct stabilization of the Doc:Coh binding interface by XMod. As shown in Fig. 3.4e, the root mean squared deviation (RMSD) of Doc increased throughout the simulation as XMod unfolded. Coh RMSD remained stable until it started to lose H-bonds with Doc. Taken together this suggests that as XMod unfolded, Coh and Doc became more mobile and lost interaction strength, potentially explaining the increase in Δx from $0.13\,\text{nm} - 0.19\,\text{nm}$ upon unfolding of XMod in the experimental datasets. Apparently the XMod is able to directly stabilize the Doc:Coh interface, presumably through contact with Doc insert sequences that then propagate this stabilizing effect to the Doc:Coh binding interface.

In summary, we investigated an ultrastable XMod-Doc:Coh complex involved in bacterial adhesion to cellulose. While previously the role of XMod functioning in tandem XMod-Doc dyads was unclear [4, 151], we show that XMod serves as a mechanical stabilizer and force-shielding effector subdomain in the ultrastable ligand-receptor complex. The Doc:Coh complex presented here exhibits one of the most mechanically robust protein-protein interactions reported thus far, and points towards new mechanically-stable artificial multi-component biocatalysts for industrial applications, including production of second-generation biofuels.

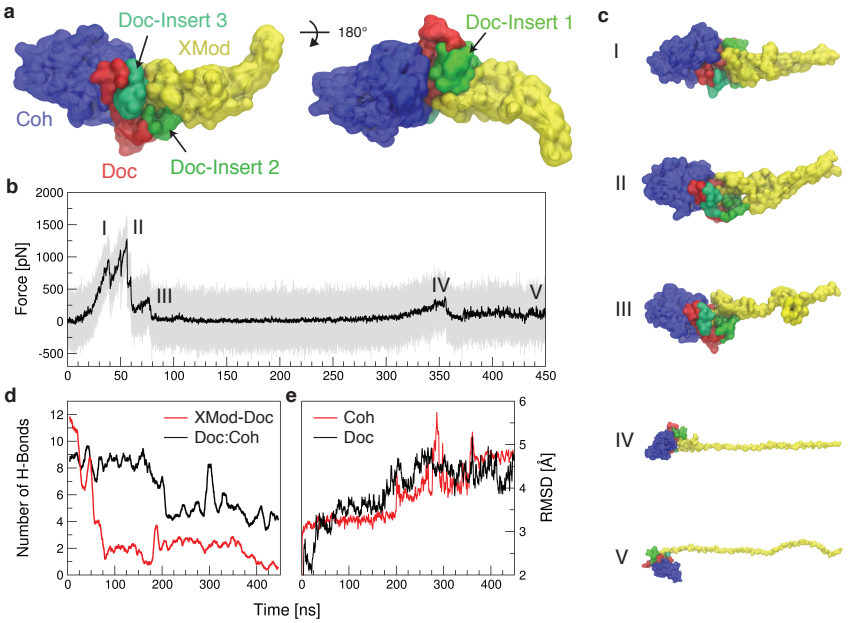


Figure 3.4: SMD shows unfolding of XMod destabilizes Doc:Coh binding interface. XMod was unfolded by moving the harmonic restraint to the C terminus of XMod while the N terminus was moved at 0.625 A ns^{-1}. (a) Surface representation of XMod-Doc:Coh complex with Doc insert sequences. Coh is shown in blue, Doc in red and green (inserts), and XMod in yellow. (b) Force time trace of XMod unfolding. The domain starts to unfold in several substeps starting at ~ 400 pN. Snapshots at different time steps are labelled I-V and are shown in (c). (d) Average number of hydrogen bonds between Doc:Coh (black) and XMod-Doc (red). XMod-Doc contact is dominated by the insert sequences 1–3. (e) Root mean squared deviation (RMSD) of Doc (black) and Coh (red).

3.3 Methods

3.3.1 Site directed mutagenesis

Site-Directed Mutagenesis of *Ruminococcus flavefaciens* strain FD1 Chimeric Cellulosomal Proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus* [4] were subjected to QuikChange mutagenesis [164] to install the following mutations: A2C in the CBM and T129C in the xylanase, respectively.
For the construction of the native configuration of the CohE-CBM A2C fusion protein Gibson Assembly [165] was used. For further analysis CohE-CBM A2C was modified with a QuikChange PCR [166] to replace the two cysteins (C2 and C63) in the protein with alanine and serine (C2A and C63S). All mutagenesis products were confirmed by DNA sequencing analysis.

The XynT6-XDoc T129C was constructed using the following primers:
5'-acaaggaaggtaagccaatggttaatgaatgcgatccagtgaaacgtgaac-3'
5'-gttcacgtttcactggatcgcattcattaaccattggcttaccttccttgt-3'

The CBM-CohE A2C was constructed using the following primers:
5'-ttaactttaagaaggagatataccatgtgcaatacaccggtatcaggcaatttgaag-3'
5'-cttcaaattgcctgataccggtgtattgcacatggtatatctccttcttaaagttaa-3'

The CohE-CBM C2A C63S was constructed using the following phosphorylated primers:
5'-ccgaatgccatggccaatacaccgg-3'
5'-cagaccttctggagtgaccatgctgc-3'

3.3.2 Expression and purification of cysteine-mutated Xyn-XMod-Doc

The T129C Xyn-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20 mM imidazole and 2 mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialyzed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50 % (v/v) glycerol at ~ 20 °C. The concentrations of the protein stock solutions were determined to be ~ 5 mg mL^{-1} by absorption spectrophotometry.

3.3.3 Expression and purification of Coh-CBM and mutated Coh-CBM C2A C63S

The Coh-CBM C2A, C63S fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicole containing ZYM-5052 media [167] overnight at 22°C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at −80°C. The concentrations of the protein stock solutions were determined to be 5 mg/mL by absorption spectrophotometry.

3.3.4 Sample preparation

Cantilevers and cover glasses were functionalized according to previously published protocols [71, 85]. Briefly, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyl-XMod-Doc proteins were reduced for 2 hours at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with TBS measurement buffer, centrifuged at 850 rcf for 3 minutes, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with TBS to

remove unbound proteins, and stored under TBS prior to force spectroscopy measurements. Site specific immobilization of the Coh-CBM-ybbR fusion proteins to previously PEGylated cantilevers or coverglasses was carried out according to previously published protocols [168]. Briefly, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer for 1h at room temperature. Levers or surfaces were then rinsed with TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 hours at room temperature. Finally, surfaces or levers were subjected to a final rinse with TBS and stored under TBS prior to measurement.

3.3.5 Single molecule force spectroscopy measurements

SMFS measurements were performed on a custom built AFM [169] controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise / equipartition method [170]. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from $0.2\,\mu m\,s^{-1} - 6.4\,\mu m\,s^{-1}$. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18-24 hours. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw datasets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single molecule interactions. We filtered the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: the xylanase ($\sim$ 89 nm) and the CBM ($\sim$ 56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

3.3.6 Data analysis

Data were analyzed using slight modifications to previously published protocols [71, 110, 154]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11\,nm$ connected by a fixed angle $\gamma = 41^\circ$ and assembled into barrier position histograms using cross-correlation. Detailed description of the contour length transformation can be found in the Supplementary Information and Supplementary Figure 3.5.

For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs. time trace in the immediate vicinity prior to the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying Gaussian fits to histograms of rupture forces and loading rates at each pulling speed.

3.3.7 Molecular dynamics simulations

The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97 A resolution and is available at the protein data bank (PDB:4IU3). A protonation analysis

performed in VMD[171] did not suggest any extra protonation and all the amino acid residues were simulated with standard protonation states. The system was then solvated, keeping also the water molecules present in the crystal structure, and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter-ions, which were randomly arranged in the solvent. Two other systems, based on the aforementioned one, were created using a similar salt concentration to the one used in the experiments (75 mM of NaCl). This additional salt caused little or no change in SMD results.

The MD simulations in the present study were performed employing the NAMD molecular dynamics package [137, 172]. The CHARMM36 force field [128, 173] along with the TIP3 water model [174] was used to describe all systems. The simulations were done assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 A was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) [175] method. The equations of motion were integrated using the r-RESPA multiple time step scheme [137] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Prior to the MD simulations all the systems were submitted to an energy minimization protocol for 1000 steps. The first two nanoseconds of the simulations served to equilibrate systems before the production runs that varied from 40 ns − 200 ns in the ten different simulations that were carried out. The equilibration step consisted of 500 ps of simulation where the protein backbone was restrained and 1.5 ns where the system was completely free and no restriction or force was applied. During the equilibration the initial temperature was set to zero and was constantly increased by 1 K every 100 MD steps until the desired temperature (300 K) was reached.

To characterize the coupling between dockerin and cohesin we performed steered molecular dynamics (SMD) simulations [120] of constant velocity stretching (SMD-CV protocol) employing 3 different pulling speeds: 1.25 A ns^{-1}, 0.625 A ns^{-1}, and 0.25 A ns^{-1}. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically (center of mass of ASN5), and moving a second restraint point, at the end of the Coh domain (center of mass of GLY210), with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Due to the flexibility of the linkers this approach reproduces the experimental setup. All analyses of MD trajectories were carried out employing VMD [171] and its plugins. Secondary structures were assigned using the Timeline plugin, which employs STRIDE criteria [176]. Hydrogen bonds were assigned based on two geometric criteria for every trajectory frame saved: first, distances between acceptor and hydrogen should be less than 3.5 A; second, the angle between hydrogen-donor-acceptor should be smaller than 30°. Surface contact areas of interacting residues were calculated employing Volarea [177] implemented in VMD. The area is calculated using a probe radius defined as an in silico rolling spherical probe that is screened around the area of the dockerin exposed to the cohesin and the also the cohesin area exposed to the dockerin.

3.4 Supplementary material

Supplementary figures

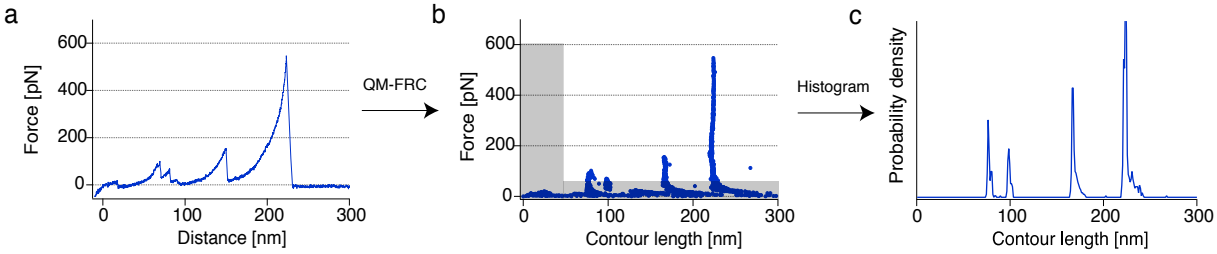


Figure 3.5: Assembly of contour length histograms. **a** Force-extension traces are transformed into contour length space using a QM-corrected FRC model with parameters $\gamma = 41°$, and $b = 0.11\,\text{nm}$. **b** In force-contour length space, force and contour length thresholds are applied and the data are histogrammed with a bin width of 1 nm to obtain the histogram in **c**. To obtain a master histogram, individual histograms reflecting a specific unfolding pathway are cross-correlated and aligned by offsetting by the maximum correlation value.

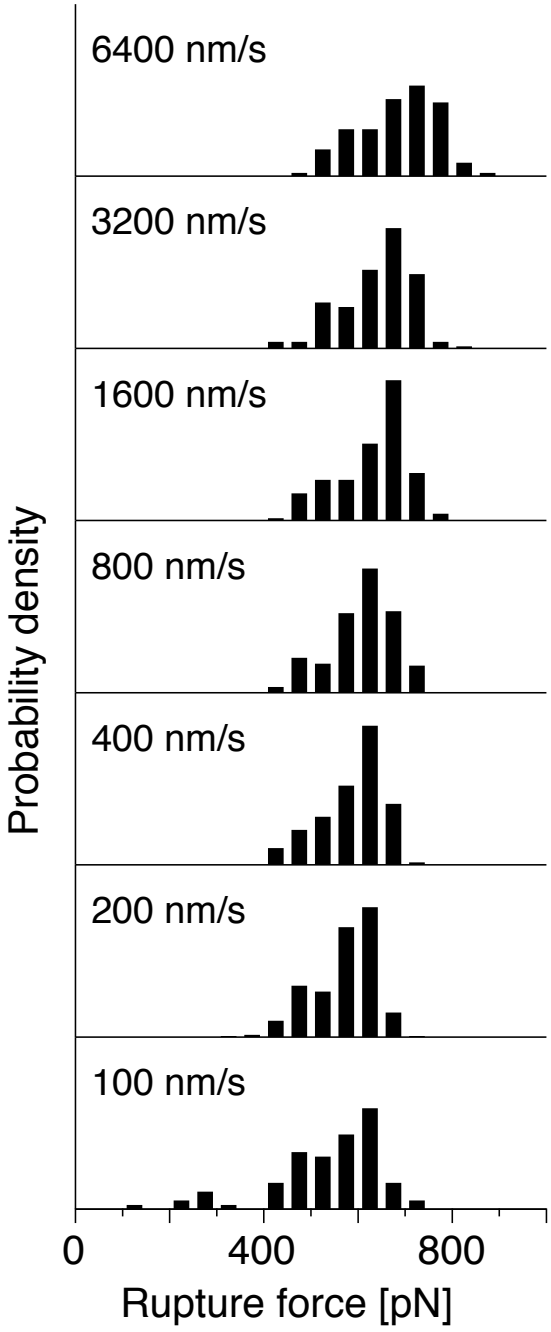


Figure 3.6: Complex rupture force histograms for pulling speeds ranging from 100 nm s^{-1} to 6400 nm s^{-1}. Pulling speeds are indicated next to the histograms. Only traces with an intact XMod were taken into account (no XMod unfolding observed, corresponding to Fig. 3.2, trace 1). At the slowest pulling speed data suggest the presence of a lower rupture force population.

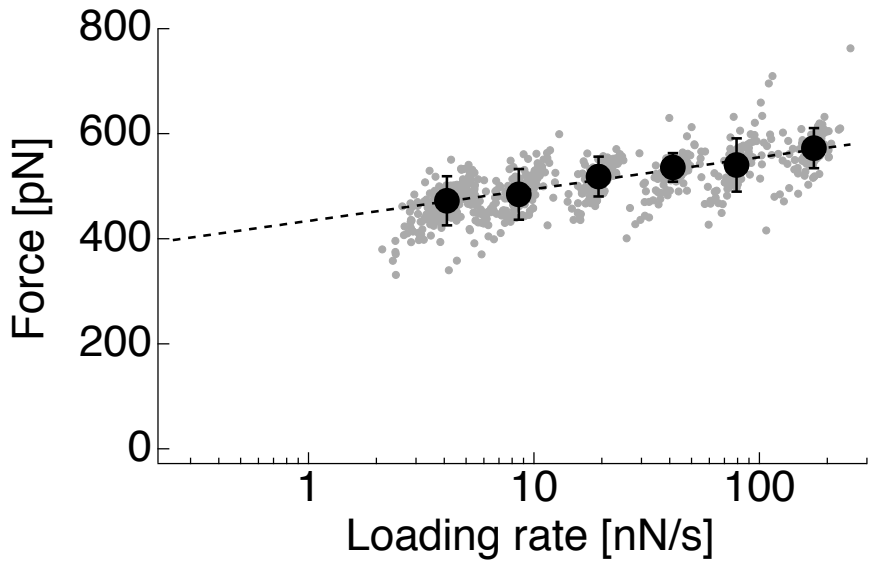


Figure 3.7: Dynamic force spectrum for XMod unfolding obtained from 654 force-extension traces. The gray points show single XMod unfolding events. Black circles represent the most probable rupture forces and loading rates obtained by Gaussian fitting at each pulling speed. Error bars are ±1 standard deviation. The dashed line is a least squares fit to the Bell-Evans model that yielded $\Delta x = 0.15\,\text{nm}$ and $k_{\text{off}} = 2.6 \times 10^{-6}\,\text{s}^{-1}$.

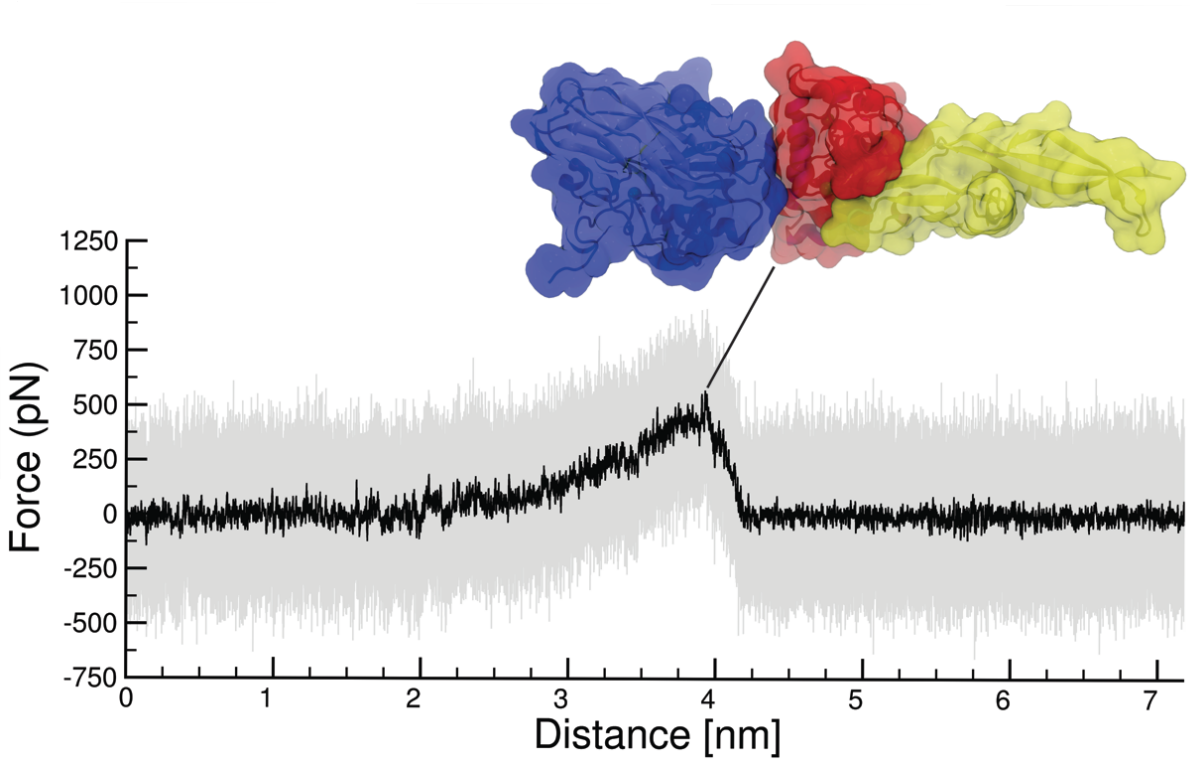


Figure 3.8: Force distance trace obtained by SMD at a pulling speed of $0.25\,\text{A ns}^{-1}$. Force values at each time step are shown in gray, with average force calculated every 200 ps in black. The inset is a snapshot of the XMod-Doc:Coh complex immediately prior to rupture. XMod is shown in yellow, Doc in red and Coh in blue.

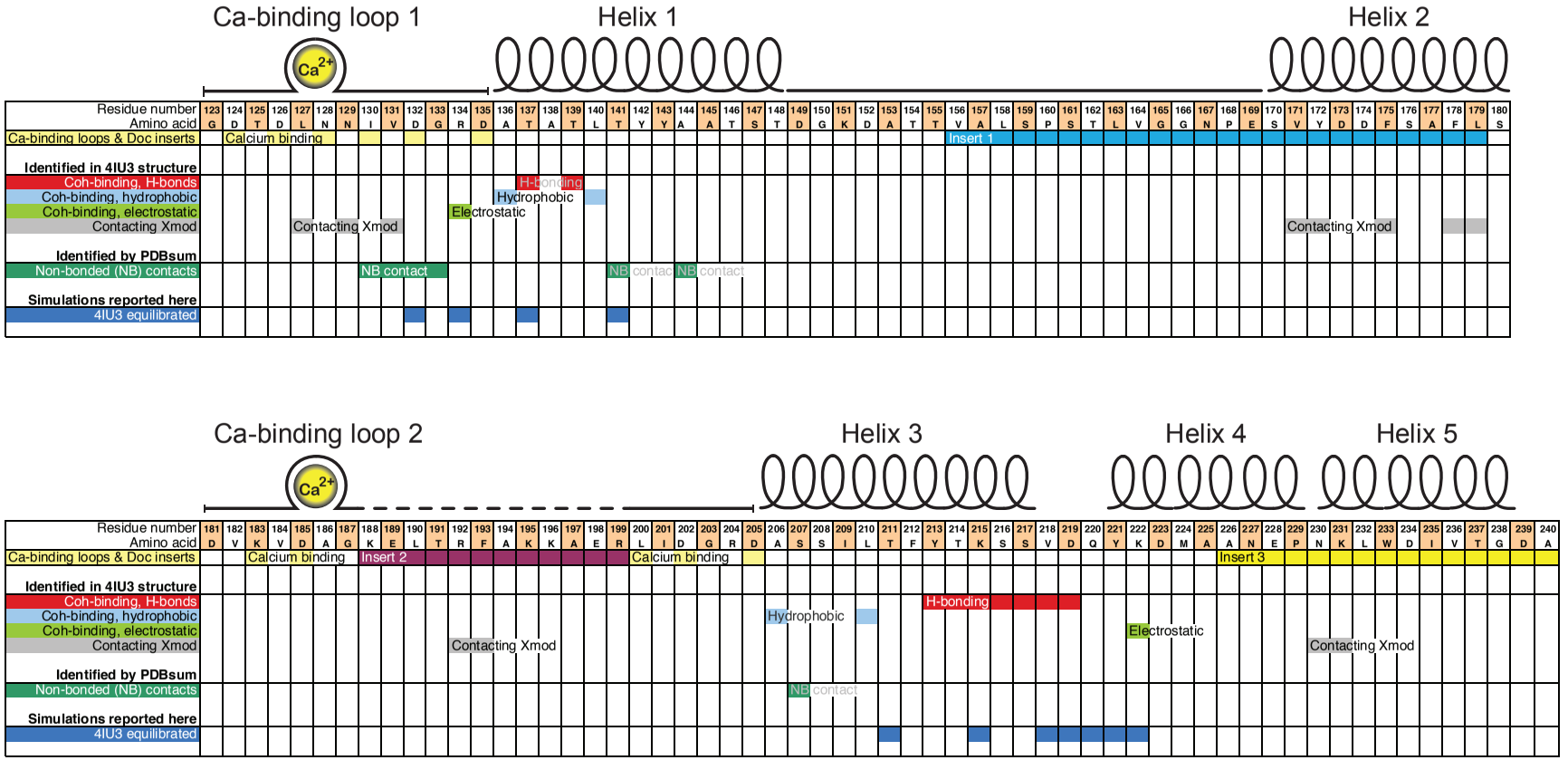
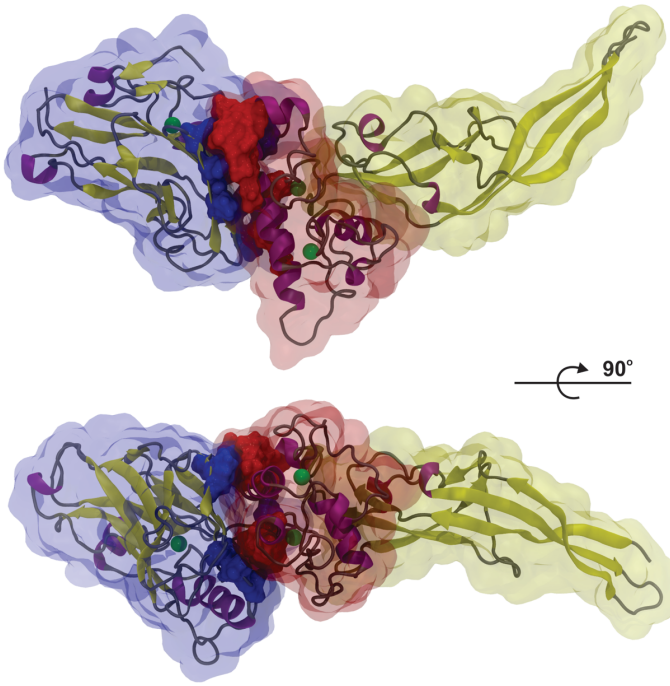


Figure 3.9: Amino acid sequence and secondary structure elements of the *Ruminococcus flavefaciens* dockerin module. The table includes interactions with the Coh and XMod domains identified in the PDB code 4IU3 crystal structure, PDBsum[178] and MD simulations.



Donor	Acceptor	Prevalence
LYS215-S	GLU105-S	87%
ASN107-S	ASP219-S	48%
ASP219-B	ASN107-S	39%
THR141-S	ASP118-S	31%
LYS156-S	ASP132-S	22%

Figure 3.10: Hydrogen bond contacts between XMod-Doc (yellow and red surface, respectively) and Coh (blue surface). The residues that have hydrogen bonds lasting for more than 10 % of the simulation time are represented in a glossy surface. In the bottom of the figure the five most prevalent hydrogen bond interactions are presented. The letter S or B indicate if the respective interaction is made by the amino acid side chain or backbone.

Supplementary tables

Module	Xylanase	CBM	X-module	Cohesin	Dockerin
No. amino acids, N_A	260 (378)	159	117	205	119
Folded length, L_F [nm]	6	2	7	2	2
Expected increment, ΔL_E [nm]	89	56	36	72	42
Observed increment, [nm]	90 ± 4	55 ± 3	34 ± 2	–	–

Table 3.1: Domain assignment of observed contour length increments. The expected contour length increment (ΔL_E) for each protein domain was calculated according to $\Delta L_E = N_A \cdot 0.365\,\text{nm} - L_F$, where L_F is the folded length, N_A is the number of amino acids, and 0.365 nm [179] is the length per stretched amino acid. L_F was measured for Xyn, CBM, and XDoc:Coh from PDB structures 1R85, 1NBC, and 4IU3, respectively. For the Xyn domain, only amino acids located C-terminal of the C129 mutation which served as attachment point are considered. Errors for the observed increments were determined from Gaussian fits to the combined contour length histogram shown in Fig. 3.2b.

Supplementary notes

Supplementary note 1: QM-FRC model for polymer elasticity

The freely rotating chain model [113] considers bonds of length b, connected by a fixed angle γ. The torsional angles are not restricted. The stretching behavior in the FRC picture is given by

$$\frac{x}{L} = \begin{cases} \frac{Fa}{3k_BT} & \text{for} \quad \frac{Fb}{k_BT} < \frac{b}{p} \\ 1 - \left(\frac{4Fp}{k_BT}\right)^{-\frac{1}{2}} & \text{for} \quad \frac{b}{p} < \frac{Fb}{k_BT} < \frac{p}{b} \\ 1 - \left(\frac{cFb}{k_BT}\right)^{-1} & \text{for} \quad \frac{p}{b} < \frac{Fb}{k_BT} \end{cases} \tag{3.1}$$

where $a = b\frac{1+\cos\gamma}{(1-\cos\gamma)\cos\frac{\gamma}{2}}$ is the Kuhn length, and $p = b\frac{\cos\frac{\gamma}{2}}{|\ln(\cos\gamma)|}$ is the effective persistence length in the FRC picture.

To account for backbone elasticity of the polypeptide chain at high force, quantum mechanical *ab-initio* calculations can be used to obtain the unloaded contour length at zero force. A polynomial approximation to these calculations can be used to obtain the unloaded contour length at zero force L_0:

$$F = \gamma_1 \left(\frac{L}{L_0} - 1\right) + \gamma_2 \left(\frac{L}{L_0} - 1\right)^2 \tag{3.2}$$

where the $\gamma_1 = 27.4\,\text{nN}$, and $\gamma_2 = 109.8\,\text{nN}$ are the elastic coefficients reported for polypeptides [117].

Supplementary note 2: Bell-Evans model for mechanically induced receptor ligand dissociation

The Bell-Evans model was used to estimate the distance to the transition state (Δx) and the natural off-rate (k_{off}) of individual rupture events:

$$\langle F \rangle = \frac{k_B T}{\Delta x} \ln \frac{\Delta x \cdot \dot{F}}{k_{off} k_B T} \tag{3.3}$$

where k_B is Boltzmann's constant, T is the temperature and $\dot{F}$ is the loading rate at the point of rupture.

Supplementary methods

Materials

Silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation) with a nominal spring constant of 100 pN/nm (25 kHz resonance frequency in water) were used. Circular coverglasses, 2.4 cm in diameter, were obtained from Menzel Gläser (Braunschweig, Germany). 3-Aminopropyl dimethyl ethoxysilane (APDMES) was purchased from ABCR GmbH (Karlsruhe, Germany). NHS-PEG-Maleimide (5 kDa) was purchased from Rapp Polymer (Tübingen, Germany). Immobilized TCEP Disulfide Reducing Gel was obtained from Thermo Scientific (Pittsburgh, PA). The following standard chemicals were obtained from Carl Roth (Karlsruhe, Germany) and used as received: tris(hydroxymethyl)aminomethane (TRIS, >99% p.a.), $CaCl_2$ (>99% p.a.), sodium borate (>99.8% p.a), NaCl (>99.5% p.a.), ethanol (>99% p.a.), and toluene (>99.5% p.a.). Borate buffer was 150 mM, pH 8.5. The measurement buffer for force spectroscopy was Tris-buffered saline (TBS, 25 mM TRIS, 75 mM NaCl, pH 7.2) supplemented with $CaCl_2$ to a final concentration of 1 mM. All buffers were filtered through a sterile 0.2 μm polyethersulfone membrane filter (Nalgene, Rochester, NY, USA) prior to use.

Protein sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

Xyn-XModDoc

Xylanase T129C
Linker or extra residues
X-module
Dockerin type III

M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q
R Y K N E F T I G A A V E P Y Q L Q N E K D V Q M L K R H F
N S I V A E N V M K P I S I Q P E E G K F N F E Q A D R I V K
F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E G K

P M V N E C D P V K R E Q N K Q L L L K R L E T H I K T I V
E R Y K D D I K Y W D V V N E V V G D D G K L R N S P W Y Q
I A G I D Y I K V A F Q A A R K Y G G D N I K L Y M N D Y N T
E V E P K R T A L Y N L V K Q L K E E G V P I D G I G H Q S
H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L
D V S M Y G W P P R A Y P T Y D A I P K Q K F L D Q A A R Y
D R L F K L Y E K L S D K I S N V T F W G I A D N H T W L D
S R A D V Y Y D A N G N V V V D P N A P Y A K V E K G K G K
D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T
S A V K T Q Y V E I E S V D G F Y F N T E D K F D T A Q I K K
A V L H T V Y N E G Y T G D D G V A V V L R E Y E S E P V D I
T A E L T F G D A T P A N T Y K A V E N K F D Y E I P V Y Y N
N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V
D G R D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T
L V G G N P E S V Y D D F S A F L S D V K V D A G K E L T R F
A K K A E R L I D G R D A S S I L T F Y T K S S V D Q Y K D M
A A N E P N K L W D I V T G D A E E E

Coh-CBM C2A, C63S

CBM (C2A, C63S)
Linker or extra residues
CohIII
ybbR-Tag

M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T
K K V F D T A A D A A G Q T V T V E F K V S G A E G K Y A T T
G Y H I Y W D E R L E V V A T K T G A Y A K K G A A L E D S
S L A K A E N N G N G V F V A S G A D D D F G A D G V M W T V
E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T
D N K D S A Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y
L V K A N A T Y A D G Y I A I K A G E P G S V V P S T Q P V T
T P P A T T K P P A T T I P P S D D P N A M A N T P V S G N L
K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I
D L S K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G
S N G S Y N G I T S N V K G T F V K M S S S T N N A D T Y L
E I S F T G G T L E P G A H V Q I Q G R F A K N D W S N Y T
Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K
E P G E L K L P R S R H H H H H H G S L E V L F Q G P D S L
E F I A S K L A

Chapter 4

Mapping mechanical force propagation through biomolecular complexes

Summary

Here we employed single-molecule force spectroscopy with an atomic force microscope (AFM) and steered molecular dynamics (SMD) simulations to reveal force propagation pathways through a mechanically ultrastable multi-domain cellulosome protein complex. We demonstrate a new combination of network-based correlation analysis supported by AFM directional pulling experiments, which allowed us to visualize stiff and soft paths through the protein complex along which force was transmitted. The results implicate specific force-propagation architectures non-parallel to the pulling axis that are advantageous for achieving high dissociation forces.

4.1 Introduction

Mechanical forces play a fundamental role in biological systems. Cells are able to sense and respond to mechanical cues in their environment by, for example, modulating gene expression patterns [180], reshaping the extracellular matrix [181], or exhibiting differential biochemical activities [182]. At the molecular level, these behaviors are governed by mechanically active proteins. Such proteins are able to sense and respond to force by undergoing conformational changes [183], exposing cryptic binding sequences [184], acting synergistically with ion channels [162], or modulating their function in a variety of ways [185–187].

Experimental methods including AFM single-molecule force spectroscopy (SMFS) allow direct measurement of molecular mechanical properties. These studies have demonstrated the importance of the shear topology involving parallel breakage of hydrogen bonds in providing mechanical stability to protein folds [19, 188]. Many globular domains and protein complexes also exhibit a directional dependence in unfolding mechanics, consisting of stiff and soft axes [23–25, 189–192]. Pulling geometry can be defined by controlling the positions of the chemical

This Chapter was published by Schoeler *et al.* [72] in *Nano Letters* and adapted with permission from the American Chemical Society.

linkages between protein monomer units through a variety of bioconjugate techniques.

Primary sequences of mechanically active proteins are extremely diverse, essentially rendering them undetectable by conventional bioinformatics approaches. Yet, another computational approach, namely molecular dynamics (MD), allow sampling of structural conformations of large and frequently mechanostable protein complexes [193, 194]. Analysis of these conformations from MD trajectories have recently led to the development of network-based correlation methods for investigating signal transmission and allosteric regulation in proteins [195–197]. In network models, local correlations of positional fluctuations in a protein are represented as a web of inter-residue connections. Within such a network, the behavior of nodes that are highly correlated and within close physical proximity can be analyzed to obtain the shortest path between two network nodes (*i.e.*, amino acids). This analysis helps to identify which connecting residues are most important for intramolecular communication [197–199]. Examination of multiple pathways, also known as suboptimal paths, within an acceptable deviation from the optimal path helps to detect the web of nodes critical for transmission of information.

Among MD methods, steered molecular dynamics (SMD) simulations, in which external forces are used to explore the response and function of proteins, have become a powerful tool especially when combined with SMFS [162]. SMD has been successfully employed in a wide range of biological systems, from the investigation of protein mechanotransduction [184, 200], to permeability of membrane channels [201, 202], and the characterization of protein-receptor interactions [5]. SMD simulations have also been used to study force propagation through proteins by employing force distribution analysis (FDA) [203, 204]. In FDA, all pair-wise forces, which are usually calculated in MD simulations, are stored in $N \times N$ matrices, where N is the number of atoms [205]. These pair-wise forces can then be used to assess a protein's response to a mechanical or allosteric signal [206]. In the FDA approach, atoms under mechanical strain are identified by subtracting forces of both loaded and unloaded states for each pair of interacting atoms [204]. However, to achieve a sufficient signal to noise ratio, FDA will often require exhaustive sampling of the conformational space [205, 207]. FDA therefore, requires more computational resources than usual SMD studies, which are frequently already computationally demanding. There is therefore a clear need for new analysis methods that enable visualization of force propagation pathways from a single SMD trajectory.

Here we implemented a novel combination of SMD, network-based correlation analysis, and thermodynamic fluctuation theory, supported by AFM-SMFS experiments to study force propagation through a protein complex subjected to different pulling geometries. We chose an ultrastable receptor-ligand interaction as a model system because of its remarkably high mechanical stability [5], which effectively improves the signal-to-noise ratio. This complex consists of two interacting protein domains called cohesin (Coh) and dockerin (Doc) that maintain bacterial adhesion of *Ruminococcus flavefaciens* (Rf) to cellulosic substrates. Doc is found within the same polypeptide chain as a stabilizing ancillary domain called X-module (XMod), located N-terminally of Doc. Based on its position with the *Rf* cellulosomal network, Coh is mechanically anchored *in vivo* at its C-terminal end to the cell surface. Our prior work demonstrated that, when force is applied to the complex in the native configuration (*i.e.*, C-terminal Coh, N-terminal XMod-Doc anchor points), the complex is extremely stable, exhibiting high rupture forces of 600 pN − 750 pN at loading rates from 1 nN s^{-1} − 100 nN s^{-1} [5]. Since the bulk equilibrium affinity of the complex is an unremarkable 20 nM [4], we hypothesized that the high mechanostability was explained by a catch bond mechanism. AFM rupture force

data and SMD simulations supported this prediction, where it was observed that the contact surface area of the two proteins increased as mechanical force was applied.

To characterize the mechanisms behind Coh:Doc high stability, here we additionally pulled the complex apart in a non-native configuration (*i.e.*, N-terminal Coh, N-terminal XMod-Doc anchor points). In the non-native pulling configuration, we found that the complex dissociated along two competing pathways with very different mechanical characteristics.

Our new dynamic network analysis protocol reveals how different mechanical behaviors are attributable to differences in the direction of force transmission across the binding interface. Together, the experiments and simulations depict a simple physical mechanism for achieving high complex rupture forces: the complex directs force along pathways orthogonal to the pulling axis.

4.2 Single molecule pulling experiments and SMD

For SMFS experiments, XMod-Doc was produced as a fusion protein with an N-terminal Xylanase (Xyn) domain. Coh was produced as either an N- or C-terminal fusion domain with a carbohydrate binding module (CBM). These fusion domains were used for site specific immobilization to a glass surface and AFM cantilever to achieve the two loading configurations shown in Fig. 4.1A and further served as marker domains with known unfolding length increments to validate single-molecule interactions and sort SMFS data traces [110].

For the native pulling configuration found *in vivo*, the CBM and XMod-Doc are loaded from their C- and N-termini, respectively (Fig. 4.1A). A representative unbinding trace for the native pulling configuration is shown in Fig. 4.1B. We measured the loading rate dependence of complex rupture using both experimental and SMD datasets (unbinding trace from SMD shown in Fig. 4.3A), and plotted them on a combined dynamic force spectrum (Fig. 4.1E). The linear Bell model produced fit parameters for the effective distance to the transition state $\Delta x = 0.13\,\text{nm}$, and the zero-force off rate $k_{\text{off}} = 7.3 \times 10^{-7}\,\text{s}^{-1}$. Both experimental and simulation data were well described by a single Bell expression, despite the differences in loading rates between experiments and simulation. This observation suggests that the application of force does not significantly change Δx for this particular configuration.

To test the influence of pulling geometry on mechanical stability, we performed SMFS and SMD on the system where Coh was pulled from the opposite terminus (*i.e.*, non-native N-terminus, *cf.* Fig. 4.1A). Unlike the native pulling geometry, this configuration exhibited two clearly distinct unbinding pathways that were characterized by different force ranges (high or low) at which the complex dissociated. We refer to these pathways as non-native high force (HF) (Fig. 4.1C) and non-native low force (LF) (Fig. 4.1D).

AFM data traces classified as non-native HF showed similar characteristics as those in the native pulling configuration (*cf.* Fig. 4.1B,C,F). The non-native LF traces, however, exhibited a markedly different unfolding behavior (Fig. 4.1D). Xyn unfolding (highlighted in orange) was regularly observed, but CBM unfolding was only very rarely observed. The complex usually did not withstand forces high enough to unfold CBM when rupturing along the non-native LF path. Among non-native LF curves, we regularly found an additional contour length increment of 17 nm – 19 nm consistent with unfolding of ~ 60 amino acids located at the N-terminus of Coh. This unfolding occurred immediately following Xyn unfolding (Fig. 4.1D, red), or alternatively prior to Xyn unfolding, or with a substep (Supporting Fig. 4.5). Taken together, it appears that

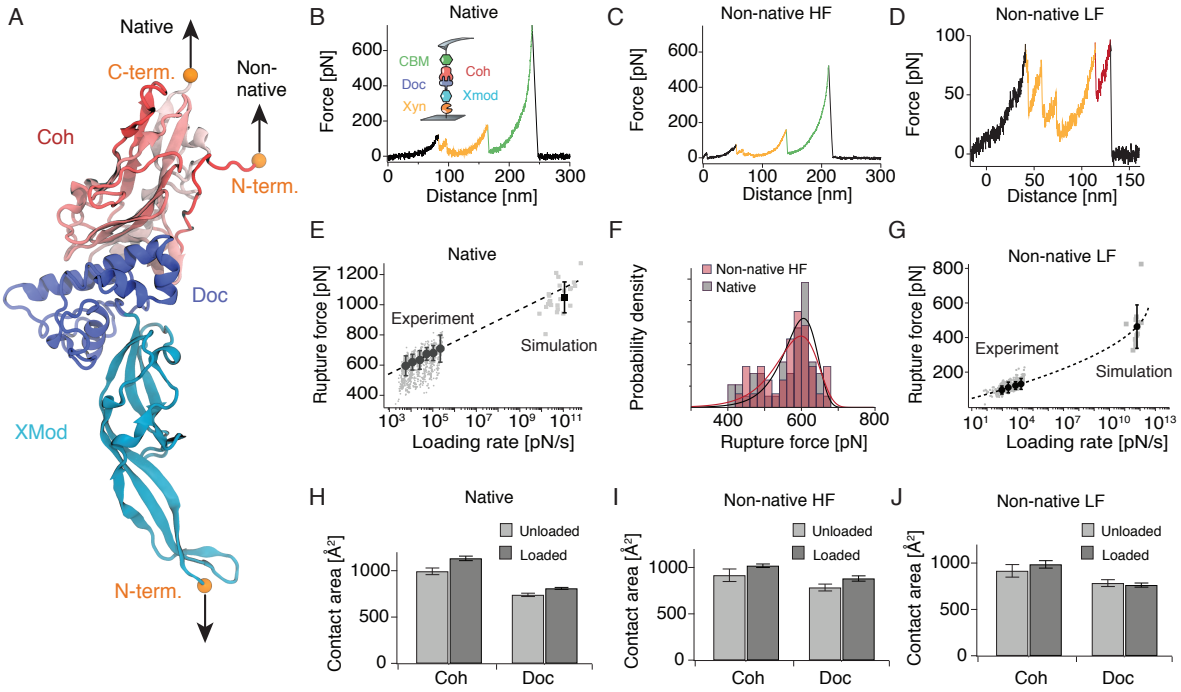


Figure 4.1: Single molecule force spectroscopy and steered molecular dynamics of XMod-Doc:Coh in two pulling configurations. (A) Crystal structure of the XMod-Doc:Coh complex (PDB 4IU3) with orange spheres marking the termini where force was applied. (B) Experimental unfolding trace for the native pulling configuration at a pulling speed of 1600 nm s^{-1}. The inset shows a schematic of the pulling geometry. Unfolding signatures of the Xyn and CBM marker domains are marked in orange and green, respectively. (C) Experimental unfolding trace for the non-native high force class obtained at a pulling speed of 700 nm s^{-1}. (D) Experimental unfolding trace for the non-native low force class obtained at a pulling speed of 700 nm s^{-1}. The additional 17 nm – 19 nm contour length increment attributed to N-terminal Coh unfolding is shown in red. (E) Dynamic force spectrum for XMod-Doc:Coh unbinding in the native geometry obtained from experiment and simulations. Gray points and squares represent the rupture force/loading rate pairs obtained from experiment and simulation, respectively. Black circles represent the most probable rupture force/loading rate obtained from Gaussian fits to the experimental data at 6 pulling speeds. The black square shows the mean rupture force and loading rate for the simulated rupture events. (F) Rupture force histograms obtained at a pulling speed of 800 nm s^{-1} for the native (gray, $n = 46$) and non-native high force class (red, $n = 48$). Fitted probability densities $p(F)$ are shown as solid black and red lines. Data for both pulling configurations were obtained with the same cantilever to minimize calibration errors. (G) Dynamic force spectrum for XMod-Doc:Coh unbinding in the non-native low force class obtained from experiments and simulation. The same representation as in (E) is used. (H, I, J) Unloaded and loaded surface contact areas for the different pulling geometries ((H) native, (I) non-native high force class, (J) non-native low force class)

partial Coh unfolding from the N-terminus destabilizes the complex, causing lower rupture forces (Fig. 4.1G).

The experimental rupture forces from the non-native HF class were indistinguishable from those arising in the native configuration. To confirm this, we performed additional measurements where both Coh configurations were alternately probed with the same Xyn-XMod-Doc functionalized cantilever (Supporting Fig. 4.6), eliminating inaccuracies introduced through multiple cantilever calibration. Most probable rupture forces at a pulling speed of 800 nm s^{-1} of 606 pN and 597 pN for the native configuration and non-native HF class respectively were determined in the Bell Evans model (Fig. 4.1F, Supporting Eq. 4.4), demonstrating that the native and non-native HF classes are experimentally indistinguishable.

For the LF class, we analyzed the final complex rupture event and plotted the combined dynamic force spectrum (Fig. 4.1G). Here, simulated and experimentally observed data were not well described by a single Bell expression. In such cases non-linear models have been developed to obtain kinetic and energetic information from dynamic force spectra [122, 123]. To fit the combined data, we used the non-linear Dudko-Hummer-Szabo (DHS) model (Supporting Eq. 4.5) and obtained values of $\Delta x = 0.42\,\text{nm}$ and $k_{\text{off}} = 0.005\,\text{s}^{-1}$. The DHS model further provides the free energy difference ΔG between the bound state and the transition state as a fit parameter, which was found to be $\Delta G = 129\,k_{\text{B}}T$. The model fit produced a distance to transition that was much longer than observed for the native configuration. Independent SMD simulations for the non-native pulling configuration were found to also lead to HF and LF unbinding scenarios (see below, Fig. 4.4A and D, respectively).

The differential solvent contact area was calculated from SMD simulations to estimate the inter-molecular contact area in the Doc:Coh complex. In the native configuration, the simulated Doc:Coh contact area increased by 14% and 9% for Coh and Doc, respectively (Fig. 4.1H). For the non-native HF class, the contact area increased by 11% and 12% for Coh and Doc, respectively (Fig. 4.1I). In the non-native LF class, the contact area increased by only 7% for Coh, and decreased by 3% for Doc (Fig. 4.1J). Evidently, an increased surface contact area for Doc in the native and non-native HF pathways correlated with high mechanostability of the system.

4.3 Force propagation theory - a simple model

To further understand the observed unbinding pathways, we sought to identify paths through the molecule along which the externally applied load propagates. From thermodynamic fluctuation theory [135, 136], it is known that the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i are related through:

$$\left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{4.1}$$

where $\Delta\mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$ and $\mathbf{r}_i$ is the position of atom i. The derivative on the right hand side of Eq. (4.1) states that neighboring atoms i and j will move with high correlation due to an external force $\mathbf{F}_i$ acting on atom i if the coupling between them is strong. Hence, a given element of a correlation matrix $M_{ij} = \left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle$ will be large in the case of a strong interaction potential between i and j. When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation, while stiff degrees become more important for the dynamics of the system.

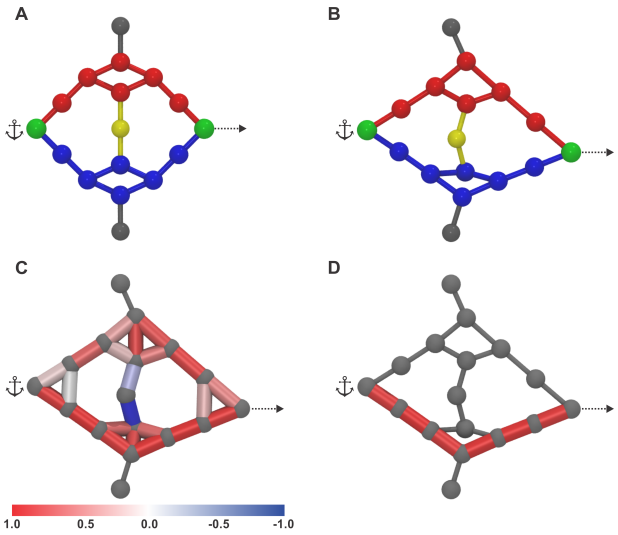


Figure 4.2: Network analysis test simulation. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red: k, blue: $5k$, yellow: $7.5k$, black: $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs where introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red.

Consequently, paths with high correlation of motion describe the paths along which force propagates through the system. To illustrate this behavior for a toy system, we employed the NAMD [137] SMD [120] constant velocity protocol to a test pattern of identical spheres connected with harmonic springs of different stiffness (Fig. 4.2A). The position of one sphere was fixed during the simulation while another sphere on the opposite side of the structure was withdrawn at constant velocity. The strained structure at the end of the simulation is shown in Fig. 4.2B. We assigned weights to the lines between spheres according to the Pearson correlation coefficient C_{ij} (Supporting Eq. 4.6) between those network nodes (Fig. 4.2C). The Pearson correlation coefficient differs from the left hand side of Eq. (4.1) by a normalization factor $\left(\left\langle \Delta\mathbf{r}_i^2(t)\right\rangle \left\langle \Delta\mathbf{r}_j^2(t)\right\rangle\right)^{-\frac{1}{2}}$ and was chosen to make our analysis mathematically more tractable. For a detailed discussion on this choice of correlation measure, see Supporting Information. In a harmonic potential approximation, the equipartition theorem can be applied to this normalization factor resulting in the following expression for C_{ij}:

$$C_{ij} = \mathrm{Tr}\,\frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \sqrt{k_{i,\mathrm{eff}} \cdot k_{j,\mathrm{eff}}} \tag{4.2}$$

where $k_{i,\mathrm{eff}} = \left(\frac{1}{k_{x_i}} + \frac{1}{k_{y_i}} + \frac{1}{k_{z_i}}\right)^{-1}$ and k_{x_i} is the curvature of the potential on atom i in the x direction. For a full derivation, see Supporting Information. Eq. (4.2) illustrates how Pearson correlation is a suitable measure to identify the stiff paths in our simple model. We then used dynamical network analysis[138] to find the path of highest correlation (Fig. 4.2D). As expected from Eq. (4.1), we found this path to be the one connected by the stiff springs.

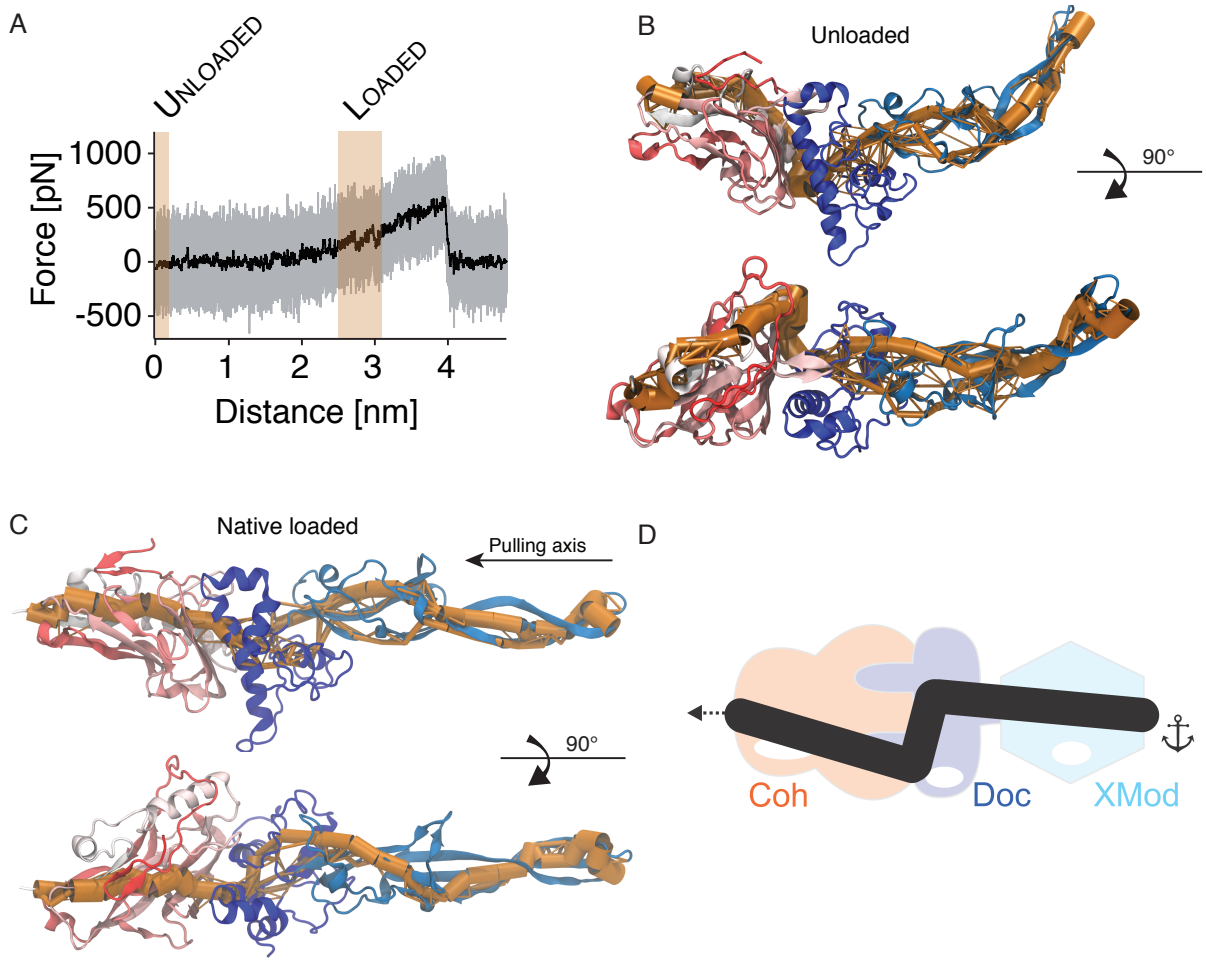


Figure 4.3: Force propagation through XMod-Doc:Coh in the native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh obtained from SMD at a pulling speed of 0.25 A ns^{-1}. The full trajectory is shown in gray. The black line represents a moving average with a box size of 500 steps. The highlighted red areas denote the windows where dynamic networks and contact areas were calculated. (B) Network paths for the unloaded system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (C) Network paths for the loaded system. A detailed 2D representation of the pathway, highlighting the amino acids present in the pathway, is shown in Fig. 4.9. (D) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large normal components to the unbinding axis.

4.4 Force propagation through XMod-Doc Coh complex

The simple pattern of spheres validated our general approach of using local correlations to identify load-bearing pathways through networks. We next employed dynamical network analysis to understand force propagation through the XMod-Doc:Coh complex.

The dynamic networks for the native configuration (unloaded and loaded) are shown in Fig. 4.3B and C, respectively. While the network shows multiple suboptimal paths in the unloaded scenario, the loaded case exhibits a well defined main path along which force propagates

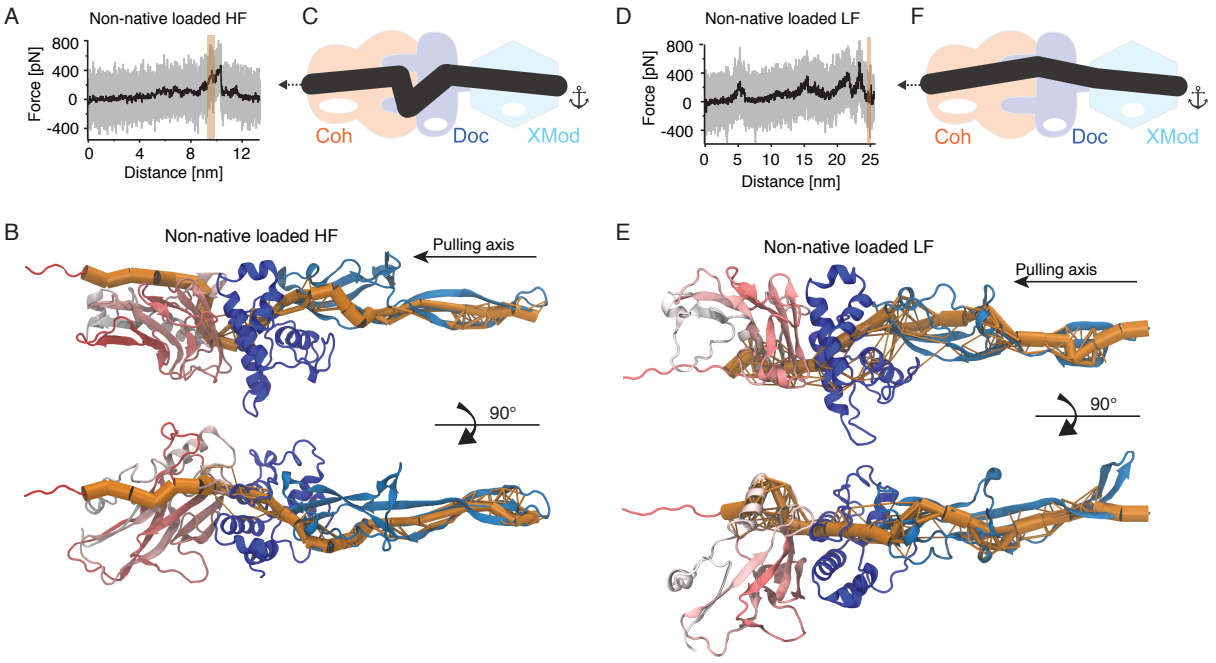


Figure 4.4: Force propagation through XMod-Doc:Coh in the non-native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 A ns^{-1}. The full trajectory is shown in gray, the black line represents a moving average with a box size of 500. Note that this computational pulling experiment revealed a high-force behavior. (B) Network pathways calculated from dynamical network analysis for the non-native HF trajectory. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supporting Fig 4.10. (C) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large normal components to the unbinding axis. (D) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 A ns^{-1}. The full trajectory is shown in gray, the black line represents a moving average with a box size of 500. This computational pulling experiment revealed partial Coh unfolding that led to LF behavior. (E) Network pathways for the non-native LF scenario. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supporting Fig. 4.11. (F) Schematic model of force propagation across the Coh:Doc binding interface. Unlike in both HF scenarios, force propagates across the binding interface mostly along the unbinding axis.

through the system. Interestingly, in the loaded configuration, force propagates through both binding helices of Doc, which results in a force path with large normal components to the unbinding axis close to the binding interface as illustrated in Fig. 4.3D. It was shown for another ultrastable protein, namely silk crystalline units, that curving force paths distribute tension through the entire system [204]. A strategy that assumes an indirect path would therefore allow the system to have more time to absorb the tension from the applied force. The result here supports the view that directing the force along a path with significant perpendicular components to the pulling axis leads to high mechanical stability. In a simple mechanical picture, a certain amount of mechanical work $\mathrm{d}W = \mathbf{F} \cdot \mathrm{d}\mathbf{s}$ is required to separate the two binding interfaces by a distance Δz and break the interaction. In this simplified picture, $\mathrm{d}\mathbf{s}$ points along the unbinding axis, whereas the force $\mathbf{F}$ is locally largely perpendicular to this direction. Consequently, a larger force is required to break the interaction than in a scenario where the force path would point along the unbinding axis.

To validate this picture, we repeated the same analysis for the non-native HF and non-

native LF pathways. The HF simulation (Fig. 4.4A) exhibited only a small stretching of the flexible N-terminal region of Coh and complex dissociation at approximately 800 pN and a pulling distance around 10 nm. However, the LF case shown in Fig. 4.4D exhibited a stepwise N-terminal Coh unfolding, dissociating at a force of about 480 pN at a pulling distance of about 25 nm. This behavior confirmed our assignment of the experimentally observed 17 nm – 19 nm contour length increment to Coh unfolding up to residue 62 in PDB 4IU3.

While the experimental data did not show a detectable difference between the native configuration and the non-native HF class, the propagation of force takes place along a different pathway (Fig. 4.4B). For N-terminal Coh pulling, helix 3 of Doc is not involved in the propagation of force as it is for the native geometry. In the native configuration, force propagated through the center of Coh, while for non-native HF the path is shifted towards the side of the molecule. Despite these differences, there is a common feature between the native and non-native HF pathways. At the binding interface, the pathway again shows pronounced normal components to the unbinding axis (cf. Fig.4.4C), suggesting that this feature is indeed responsible for the exceptional mechanical strength observed for these two unbinding pathways.

Fig. 4.4E shows the force propagation pathway for the non-native LF class prior to rupture. Due to the unfolding of the N-terminal Coh segment, the propagation of force is shifted even further away from the central portion of Coh than for the non-native HF class. Interestingly, force is propagated through the small helical segment of Coh (ALA167-GLN179), a portion of the molecule that is not involved in force propagation for any of the other analyzed trajectories. Unlike in the aforementioned scenarios, there is no pronounced tendency for normal force components at the binding interface for the non-native LF class. In fact, the force is propagated along a path largely parallel to the pulling axis (cf. Fig. 4.4F). In cases where force propagation occurs parallel to the pulling axis, as in Fig. 4.4E, low mechanical stability was observed.

The aforementioned force propagation architecture along with the effect of increasing contact surface area upon mechanical loading combine for elevated mechanostability of the system. In cases where we observed an N-terminal Coh unfolding of 62 amino acids in the non-native geometry, the system was no longer able to summon this mechanism, causing dissociation at much lower forces.

Previously, our groups have reported on a family of mechanically stable protein ligand receptor complexes that are key building blocks of cellulosomes [5, 70, 71, 154], the multi-enzyme complexes used by select anaerobic bacteria to digest lignocellulose. However, the molecular origins of the stability of these complexes remained largely unclear. An initial clue was obtained when, in a previous work, we were able to show that contact surface area of the two proteins increased as mechanical force was applied [5]. In a different study [208], coarse-grained MD simulations showed much smaller rupture forces at similar loading rates both for native and non-native pulling than we report here. This disagreement is likely due to the inability of the coarse-grained model to capture the rearrangement of amino acid side chains observed here. As we demonstrated, force propagation calculation from network-based correlation analysis helped in investigating the dramatic effect on the mechanical stability of the Doc:Coh interaction when different pulling geometries are applied. Our methodological approach, to the best of our knowledge, has never been applied even though network analysis of SMD trajectories was performed before to probe the mechanism of allosteric regulation in imidazole glycerol phosphate synthase [209].

4.5 Conclusion

In summary, for both unbinding cases where we observed high mechanostability, we found that across the binding interface, force propagated along paths with strong normal components to the pulling direction. Such a behavior was not observed for the non-native LF class, where, presumably due to N-terminal Coh unfolding, the system was no longer able to direct the force across the binding interface at high angles. From these findings, we conclude that the ultrastable complex formed by Coh and Doc achieves its remarkable mechanostability by actively directing an externally applied force toward an unfavorable angle of attack at the binding interface, consequently requiring more force to achieve a given amount of separation along the pulling direction. Our results show that this mechanically stable complex uses an architecture that exploits simple geometrical and physical concepts from Newtonian mechanics to achieve high stability against external forces. The analytical framework derived here provides a basis for developing a deeper understanding of the functioning of various mechanoactive proteins that are crucial for physiologically relevant processes such as mechanotransduction, cellular mechanosensing, and pathogenesis. Additionally, it could provide a design platform for development of artificial mechanoactive systems with applications as tissue engineering scaffolds or components in engineered nanomaterials.

4.6 Materials and methods

4.6.1 Site directed mutagenesis

We performed site-directed mutagenesis of *Ruminococcus flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus* [4] were subjected to QuikChange mutagenesis to install the mutations described in the prior paper [5]. All mutagenesis products were confirmed by DNA sequencing analysis.

4.6.2 Expression and purification of cysteine mutated Xyn-XMod-Doc

The Xyn(T129C)-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20 mM imidazole and 2 mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialyzed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at $\sim$ 20 °C. The concentrations of the protein stock solutions were determined to be $\sim$ 5 mg mL^{-1} by absorption spectrophotometry.

4.6.3 Expression and purification of Coh-CBM and mutated Coh-CBM C63S

The Coh-CBM (C63S) fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicole containing ZYM-5052 media [167] overnight at 22 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at −80 °C. The concentrations of the protein stock solutions were determined to be $\sim 5\,\mathrm{mg\,mL^{-1}}$ mg/mL by absorption spectrophotometry.

4.6.4 Sample preparation

Cantilevers and cover glasses were functionalized according to previously published protocols [71]. Briefly, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyl-XMod-Doc proteins were reduced for 2 hours at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with TBS measurement buffer, centrifuged at 850 rcf for 3 minutes, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with TBS to remove unbound proteins, and stored under TBS prior to force spectroscopy measurements. Site specific immobilization of the Coh-CBM-ybbR fusion proteins to PEGylated cantilevers or coverglasses was carried out according to previously published protocols [168]. Briefly, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer for 1h at room temperature. Levers or surfaces were then rinsed with TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 hours at room temperature. Finally, surfaces or levers were subjected to a final rinse with TBS and stored under TBS prior to measurement.

4.6.5 Single molecule force spectroscopy measurements

SMFS measurements were performed on a custom built AFM controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise / equipartition method. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from $0.2\,\mu\mathrm{m\,s^{-1}} - 6.4\,\mu\mathrm{m\,s^{-1}}$. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18-24 hours. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw datasets contained a majority of unusable curves

due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single molecule interactions with CBM and Xyn unfolding length increments. We sorted the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: the xylanase (~ 89 nm) and the CBM (~ 56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

4.6.6 Data analysis

Data were analyzed using slight modifications to previously published protocols [71, 110, 154]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11$ nm connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs. time trace in the immediate vicinity prior to the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying probability density fits to histograms of rupture forces and loading rates at each pulling speed.

4.6.7 Molecular dynamics simulations

Connecting dynamics to structural data from diverse experimental sources, molecular dynamics simulations allow one to explore off-equilibrium properties of protein structure complexes in unparalleled detail [193]. More specifically, molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes [194]. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97 A resolution and is available at the protein data bank (PDB:4IU3). The system was then solvated and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter-ions, which were randomly arranged in the solvent. Total system size was approximately 580k atoms. The MD simulations in the present study were performed employing the molecular dynamics package NAMD [137, 172]. The CHARMM36 force field [128, 173] along with the TIP3 water model [174] was used to describe all systems. The simulations were carried out assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 A was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME)[175] method. The equations of motion were integrated using the r-RESPA multiple time step scheme [137] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. The first two nanoseconds of the simulations served to equilibrate systems before the production runs, which varied from $0.2\,\mu s - 1.3\,\mu s$ in the different simulations. To characterize the coupling between dockerin and cohesin, we performed SMD simulations [120] of constant velocity stretching (SMD-CV protocol) with pulling speed of $0.25\,\text{A}\,\text{ns}^{-1}$. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically, and moving a second restraint point, at the end of the Coh domain, with constant velocity in the

desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic pulling spring is then monitored during the time of the molecular dynamics simulation. All analyses of MD trajectories were carried out employing VMD [171] and its plugins. Surface contact areas of interacting residues were calculated employing Volarea [177] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling sphere that is scanned around the area of the dockerin exposed to the cohesin and also the cohesin area exposed to the dockerin. The Network View plugin [210] on VMD [171] was employed to perform dynamical network analysis. A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 A of each other for at least 75% of the frames analyzed. As suggested by Sethi et al. [138], nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as $w_{ij} = -log\left(\mid C_{ij} \mid\right)$, where C_{ij} is the correlation matrix calculated with Carma [211]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was $-log\,(0.5) = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently.

4.7 Protein sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

4.7.1 HIS-Xyn(T128C)-XDoc

X-module
Dockerin type III
Xylanase
Linker or extra residues

M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q
R Y K N E F T I G A A V E P Y Q L Q N E K D V Q M L K R H F
N S I V A E N V M K P I S I Q P E E G K F N F E Q A D R I V K
F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E G K
P M V N E **C** D P V K R E Q N K Q L L L K R L E T H I K T I V
E R Y K D D I K Y W D V V N E V V G D D G K L R N S P W Y Q
I A G I D Y I K V A F Q A A R K Y G G D N I K L Y M N D Y N T
E V E P K R T A L Y N L V K Q L K E E G V P I D G I G H Q S
H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L
D V S M Y G W P P R A Y P T Y D A I P K Q K F L D Q A A R Y
D R L F K L Y E K L S D K I S N V T F W G I A D N H T W L D

S R A D V Y Y D A N G N V V V D P N A P Y A K V E K G K G K
D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T
S A V K T Q Y V E I E S V D G F Y F N T E D K F D T A Q I K K
A V L H T V Y N E G Y T G D D G V A V V L R E Y E S E P V D I
T A E L T F G D A T P A N T Y K A V E N K F D Y E I P V Y Y N
N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V
D G R D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T
L V G G N P E S V Y D D F S A F L S D V K V D A G K E L T R F
A K K A E R L I D G R D A S S I L T F Y T K S S V D Q Y K D M
A A N E P N K L W D I V T G D A E E E

4.7.2 Coh-CBM(C2A,C63S)-HIS-ybbR

CohIII
CBM (C2A, C63S)
ybbR-Tag
Linker or extra residues

M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T
K K V F D T A A D A A G Q T V T V E F K V S G A E G K Y A T T
G Y H I Y W D E R L E V V A T K T G A Y A K K G A A L E D S
S L A K A E N N G N G V F V A S G A D D D F G A D G V M W T V
E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T
D N K D S A Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y
L V K A N A T Y A D G Y I A I K A G E P G S V V P S T Q P V T
T P P A T T K P P A T T I P P S D D P N A M A N T P V S G N L
K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I
D L S K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G
S N G S Y N G I T S N V K G T F V K M S S S T N N A D T Y L
E I S F T G G T L E P G A H V Q I Q G R F A K N D W S N Y T
Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K
E P G E L K L P R S R H H H H H H G S L E V L F Q G P D S L
E F I A S K L A

4.7.3 CBM(T2C)-Coh-HIS

CBM (T2C)
CohIII
Linker or extra residues

M **C** N T P V S G N L K V E F Y N S N P S D T T N S I N P Q F
K V T N T G S S A I D L S K L T L R Y Y Y T V D G Q K D Q T F
W C D H A A I I G S N G S Y N G I T S N V K G T F V K M S S S
T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R F A K
N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L

N G V L V W G K E P G G S V V P S T Q P V T T P P A T T K P
P A T T I P P S D D P N A M A L T D R G M T Y D L D P K D G
S S A A T K P V L E V T K K V F D T A A D A A G Q T V T V E F
K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G A
Y A K K G A A L E D S S L A K A E N N G N G V F V A S G A D D
D F G A D G V M W T V E L K V P A D A K A G D V Y P I D V A Y
Q W D P S K G D L F T D N K D S A Q G K L M Q A Y F F T Q G
I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I K A G
E P L E H H H H H H

4.8 Supplementary material

The Pearson correlation matrices of the Xmod-Doc:Coh complex before and after applying force in the native pulling configuration are presented in Supplementary Figure S3 and S4, respectively. For the unloaded complex, movements within Doc domain are seen to be highly correlated, while XMod is seen to be divided into two anti-correlated sub-domains, one comprising the β-sheet fragment close to the N-terminus (residues 5-15 and 45-66) and the other constituting the rest of the domain. Intra-domain correlations of Coh exhibit more a complex pattern to which both secondary (anti-parallel β-strands and β-sheet at the binding interface) and tertiary structure (vicinity of C- and N-termini) contribute. Some of the inter-domain correlations in the complex originate from spatial vicinity and direct interactions, specifically at the Doc:Coh binding interface and at XMod contacts with Doc inserts. However, coupling between distant parts of the complex is also present. For example, fluctuations of the non-binding part of Coh are correlated with the N-terminal part of XMod and strongly anti-correlated with Doc domain.

4.8.1 Constant barrier distance model

The constant barrier distance model [120], also referred to as the Bell-Evans model [212], is commonly used to estimate the distance to the transition state Δx and the natural off-rate k_0 of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments. It predicts that the most probable rupture force $\langle F \rangle$ is linearly dependent on the logarithm of the force loading rate [120]:

$$\langle F(r) \rangle = \frac{k_B T}{\Delta x} \ln \frac{\Delta x \cdot r}{k_0 k_B T} \tag{4.3}$$

where k_B is Boltzmann's constant, T is the temperature and r is the loading rate at the point of rupture.

The probability density distribution of rupture forces at given loading rate r in this model is given as[120]:

$$p(F) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 \cdot k_B T}{\Delta x \cdot r}\left(e^{\frac{\Delta x}{k_B T} F} - 1\right)\right] \tag{4.4}$$

4.8.2 Dudko-Hummer-Szabo model

The Dudko-Hummer-Szabo (DHS)[123, 213] model describes a non-linear dependence for the most probable rupture force on loading rate:

$$\langle F(r)\rangle = \frac{\Delta G}{\nu \Delta x}\left\{1 - \left[\frac{k_B T}{\Delta G}\ln\left(\frac{k_B T k_0}{\Delta x r} e^{\frac{\Delta G}{k_B T}+\gamma}\right)\right]^{\nu}\right\} \tag{4.5}$$

where ΔG is the free energy of activation and $\gamma = 0.577$ is the Euler-Mascheroni constant. The model parameter ν defines the single-well free-energy surface model used ($\nu = \frac{2}{3}$ for linear-cubic and $\frac{1}{2}$ for cusp free-energy. For $\nu = 1$ and $\Delta G \to \infty$ independent of ν the Eqs. (4.3) and (4.4) are recovered.

4.8.3 Pearson correlation and covariance matrix

Validation

An $N \times N$ matrix of Pearson correlation coefficients C_{ij} (Supporting Eq. S4) was calculated from each atom's x, y, z position throughout the simulation trajectory, which inherently ignores off-diagonal elements of the atomic 3×3 submatrices D_{ij}^{mn} from the full normalized $3N \times 3N$ covariance matrix (*i.e.*, correlations along orthogonal axes are neglected, see Supporting Eqs. (4.7) and (4.8)) and Supporting Fig 4.12.

Although this quasi-harmonic approximation is commonly employed in correlation analysis [198, 209, 210, 214–216], it is not *a priori* justified for complicated biomolecular interactions [217]. To validate the use of Pearson correlations, we therefore first analyzed independently the contributions from diagonal and off-diagonal elements of each 3 x 3 covariance submatrix for each pair of α-carbons within the structure (Fig. 4.13A and B). Both with and without applied force, the off-diagonal elements roughly follow Gaussian distributions centered around a correlation value of 0. Interestingly, as force was applied, the standard deviation of the distribution of off-diagonal correlation values decreased from $\sigma_{unloaded} = 0.45$ to $\sigma_{loaded} = 0.29$. This indicated a lesser influence of off-diagonal elements on the highly (anti-)correlated motion within the system under force (see Supporting Discussion 3). The diagonal elements of the sub-matrices that are used for calculating the Pearson correlation values showed a dramatically different behavior. Both in the unloaded and loaded state, the resulting distributions were strongly shifted towards highly correlated motion, and the shape of the distribution remained mostly unchanged after application of force. Since our analysis relies on the identification of paths of highest correlation through proximate residues, the quasi-harmonic approximation implied by the use of Pearson correlation is justified, especially for suboptimal pathway analysis. The resulting distributions of on- and off-diagonal matrix elements of each covariance submatrix for the loaded configuration HF class (Fig. 4.14A) and LF class (Fig. 4.14B) exhibited the same characteristics as previously described for the native configuration, with off-diagonal elements showing symmetric correlations around zero and diagonal elements showing highly correlated motions.

Supplementary equations

The Pearson correlation coefficient C_{ij} used in our dynamical network analysis protocol is given by:

$$C_{ij} = \frac{\left\langle \Delta \mathbf{r}_i(t) \cdot \Delta \mathbf{r}_j(t) \right\rangle}{\left(\left\langle \Delta \mathbf{r}_i(t)^2 \right\rangle \left\langle \Delta \mathbf{r}_j(t)^2 \right\rangle \right)^{\frac{1}{2}}} \tag{4.6}$$

where $\Delta \mathbf{r}_i(t) = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$.

The full $3N \times 3N$ covariance matrix M_{ij} for atoms i and j consists of 3×3 submatrices of the form:

$$\left\langle \Delta \mathbf{r}_i(t) \, \Delta \mathbf{r}_j(t)^T \right\rangle = M_{ij} = \begin{pmatrix} M_{ij}^{xx} & M_{ij}^{xy} & M_{ij}^{xz} \\ M_{ij}^{yx} & M_{ij}^{yy} & M_{ij}^{yz} \\ M_{ij}^{zx} & M_{ij}^{zy} & M_{ij}^{zz} \end{pmatrix} \tag{4.7}$$

The full normalized correlation matrix is calculated from M_{ij}:

$$D_{ij}^{mn} = \frac{M_{ij}^{mn}}{\sqrt{M_{ij}^{mm} M_{ij}^{nn}}} \tag{4.8}$$

Consequently, the Pearson correlation coefficient is calculated as the trace of the normalized 3×3 submatrices ($C_{ij} = \mathrm{Tr}\, D_{ij}$).

Derivation of main text equation (4.1)

Eq. 4.1 from the main text reads:

$$\left\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \right\rangle = k_\mathrm{B} T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{4.9}$$

Eqs. (4.1) and (4.6) can be combined:

$$C_{ij} = k_\mathrm{B} T \, \mathrm{Tr} \, \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\left\langle \Delta \mathbf{r}_i^2(t) \right\rangle \left\langle \Delta \mathbf{r}_j^2(t) \right\rangle \right)^{-\frac{1}{2}} \tag{4.10}$$

For an arbitrary potential $U_i(\mathbf{r})$ of atom i, a Taylor expansion around the potential minimum (set to be at 0) yields:

$$U_i(\mathbf{r}) = 0 + \underbrace{\mathbf{r}_i^T \nabla U(0)}_{=0} + \frac{1}{2} \mathbf{r}_i^T \underline{H}(0) \, \mathbf{r}_i + \ldots \tag{4.11}$$

where $\underline{H}(0)$ is the Hessian matrix evaluated at the potential minimum. Assuming Schwarz' theorem holds for $U_i(\mathbf{r})$, $\underline{H}(0)$ is a symmetric matrix and therefore has real eigenvalues and orthonormal eigenvectors. Hence, a change to the eigenbasis of $\underline{H}(0)$ is a rotation of the coordinate system. In this new basis the Hessian is diagonal:

$$\underline{H}(0) \rightarrow \underline{H}'(0) = \begin{pmatrix} k_{x'} & 0 & 0 \\ 0 & k_{y'} & 0 \\ 0 & 0 & k_{z'} \end{pmatrix} \tag{4.12}$$

This yields a simple expression for the second order term in Eq. (4.11):

$$U_i\left(\mathbf{r}'\right) = \frac{1}{2}\mathbf{r}'^T \underline{H}'\left(0\right)\mathbf{r}' = \frac{1}{2}\left(k_{x'}x'^2 + k_{y'}y'^2 + k_{z'}z'^2\right) \tag{4.13}$$

Now we inspect the normalization of C_{ij}:

$$\langle \Delta\mathbf{r}_i^2\left(t\right)\rangle = \langle \mathbf{r}_i^2\left(t\right) - 2\mathbf{r}_i\left(t\right)\langle \mathbf{r}_i\left(t\right)\rangle + \langle \mathbf{r}_i\left(t\right)\rangle^2\rangle \tag{4.14}$$

In the harmonic approximation of the potential of atom i, $\langle \mathbf{r}_i\left(t\right)\rangle = 0$, and therefore $\langle \Delta\mathbf{r}_i\left(t\right)^2\rangle = \langle \mathbf{r}_i^2\left(t\right)\rangle$. In the basis of $\underline{H}'\left(0\right)$ this becomes:

$$\langle \mathbf{r}_i'^2\left(t\right)\rangle = \langle x_i'\left(t\right)^2 + y_i'\left(t\right)^2 + z_i'\left(t\right)^2\rangle = \langle x_i'\left(t\right)^2\rangle + \langle y_i'\left(t\right)^2\rangle + \langle z_i'\left(t\right)^2\rangle \tag{4.15}$$

Applying the equipartition theorem to this result yields:

$$\langle x_i'\left(t\right)^2\rangle = \frac{k_\mathrm{B}T}{k'_{x_i}} \tag{4.16}$$

And therefore:

$$\langle \Delta\mathbf{r}_i'\left(t\right)^2\rangle = k_\mathrm{B}T\left(\frac{1}{k'_{x_i}} + \frac{1}{k'_{y_i}} + \frac{1}{k'_{z_i}}\right) = \frac{k_\mathrm{B}T}{k'_{i,\mathrm{eff}}} \tag{4.17}$$

Plugging this result into Eq. (4.10), one finds:

$$C_{ij} = k_\mathrm{B}T\,\mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\left(\frac{k_\mathrm{B}T}{k'_{i,\mathrm{eff}}}\right)^{-\frac{1}{2}}\left(\langle \Delta\mathbf{r}_j\left(t\right)^2\rangle\right)^{-\frac{1}{2}} \tag{4.18}$$

Repeating the above steps for atom j yields the final result:

$$C_{ij} = k_\mathrm{B}T\,\mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\left(\frac{k_BT}{k'_{i,\mathrm{eff}}}\right)^{-\frac{1}{2}}\left(\frac{k_BT}{k'_{j,\mathrm{eff}}}\right)^{-\frac{1}{2}} \tag{4.19}$$

$$= \mathrm{Tr}\,\frac{\partial\mathbf{r}_j}{\partial\mathbf{F}_i}\cdot\sqrt{k'_{i,\mathrm{eff}}\cdot k'_{j,\mathrm{eff}}} \tag{4.20}$$

4.9 Supplementary figures

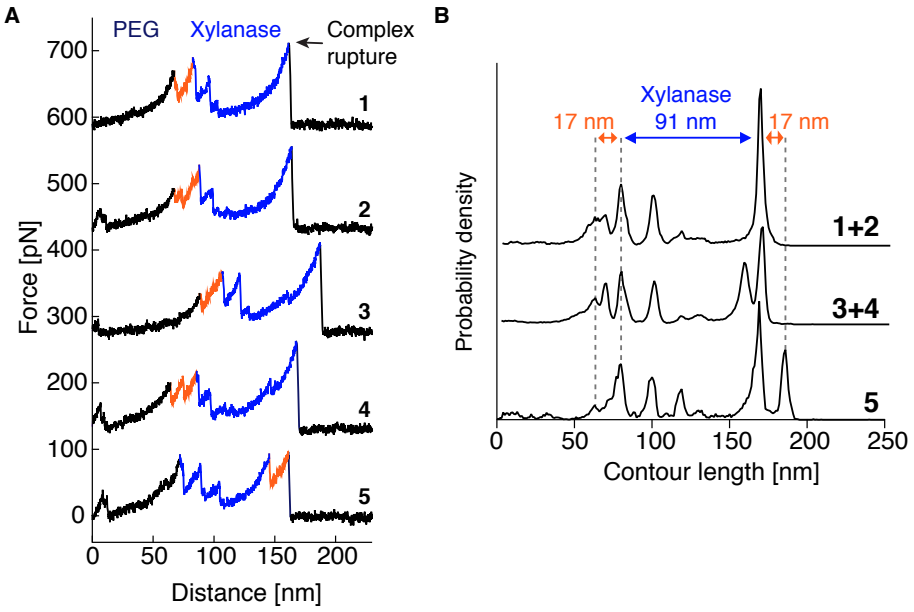


Figure 4.5: SMFS of the non-native low force curve class. **A** Typical unfolding fingerprints. All traces showed a characteristic Xyn fingerprint (blue). A 17 nm – 19 nm increment corresponding to partial N-terminal Coh unfolding (orange) occurs either prior to Xyn unfolding (traces 1-4), or just before complex rupture (trace 5). It was observed as a single event (traces 1,3 and 5) or showed substructure (traces 2 and 4). **B** Traces were grouped and assembled into contour length histograms. One or more of the unassigned increments combined into a 17 nm – 19 nm increment.

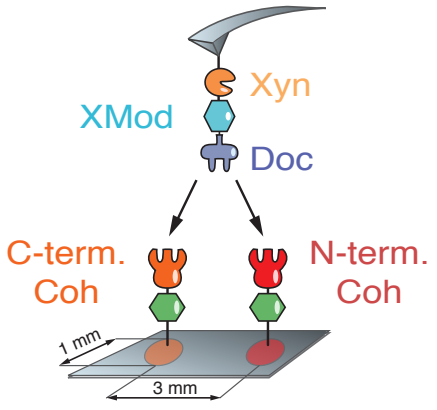


Figure 4.6: Comparing the native geometry with the non-native high force class. Two exclude uncertainties in cantilever calibration when comparing the native geometry with the non-native HF class, we immobilized both Coh-CBM (native) and CBM-Coh (non-native) on two spatially separated spots on a single cover glass. These spots where then alternately probed with the same Xyn-XMod-Doc functionalized cantilever.

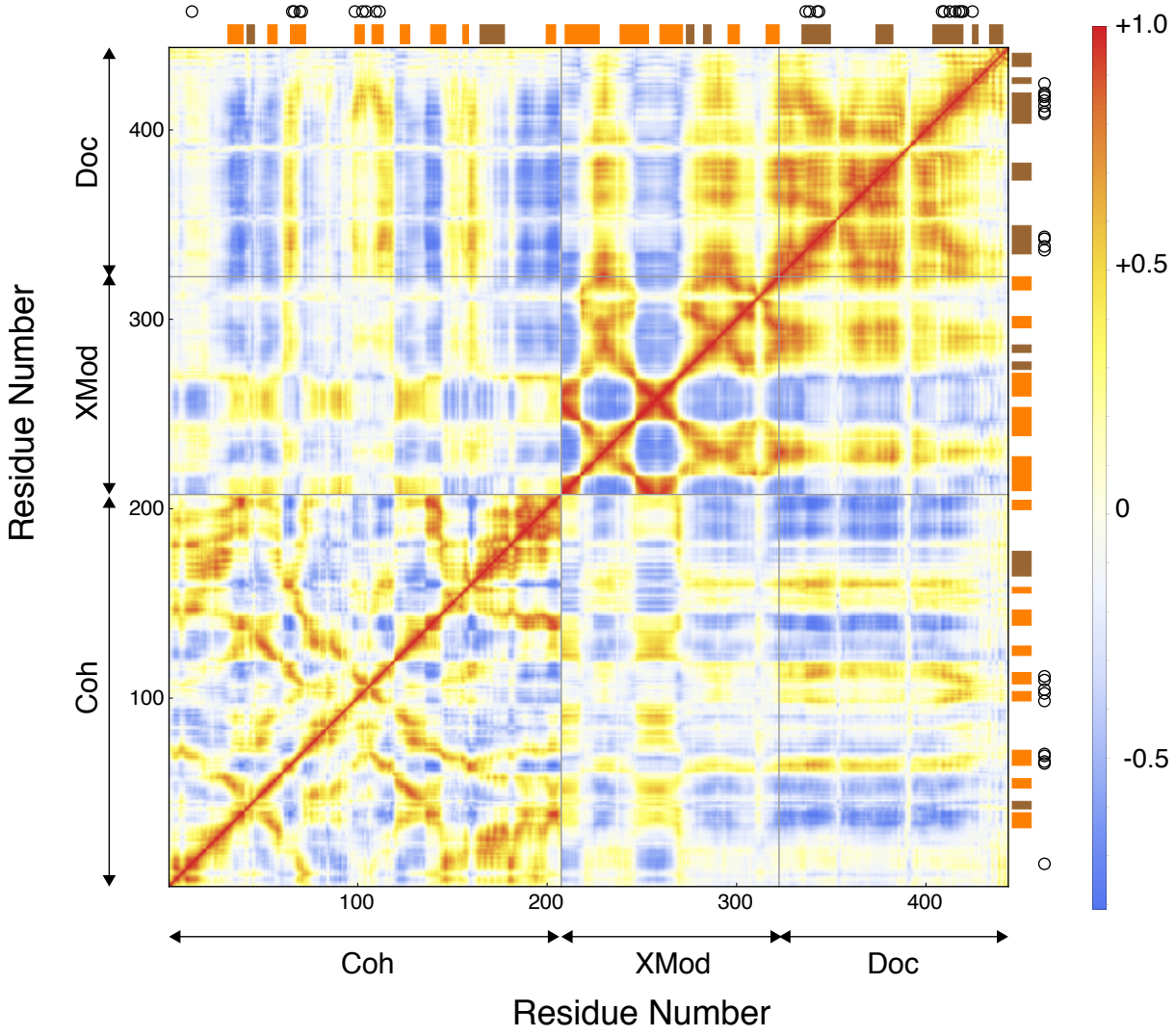


Figure 4.7: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the unloaded Xmod-Doc:Coh complex. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from the Coh and Doc binding interface.

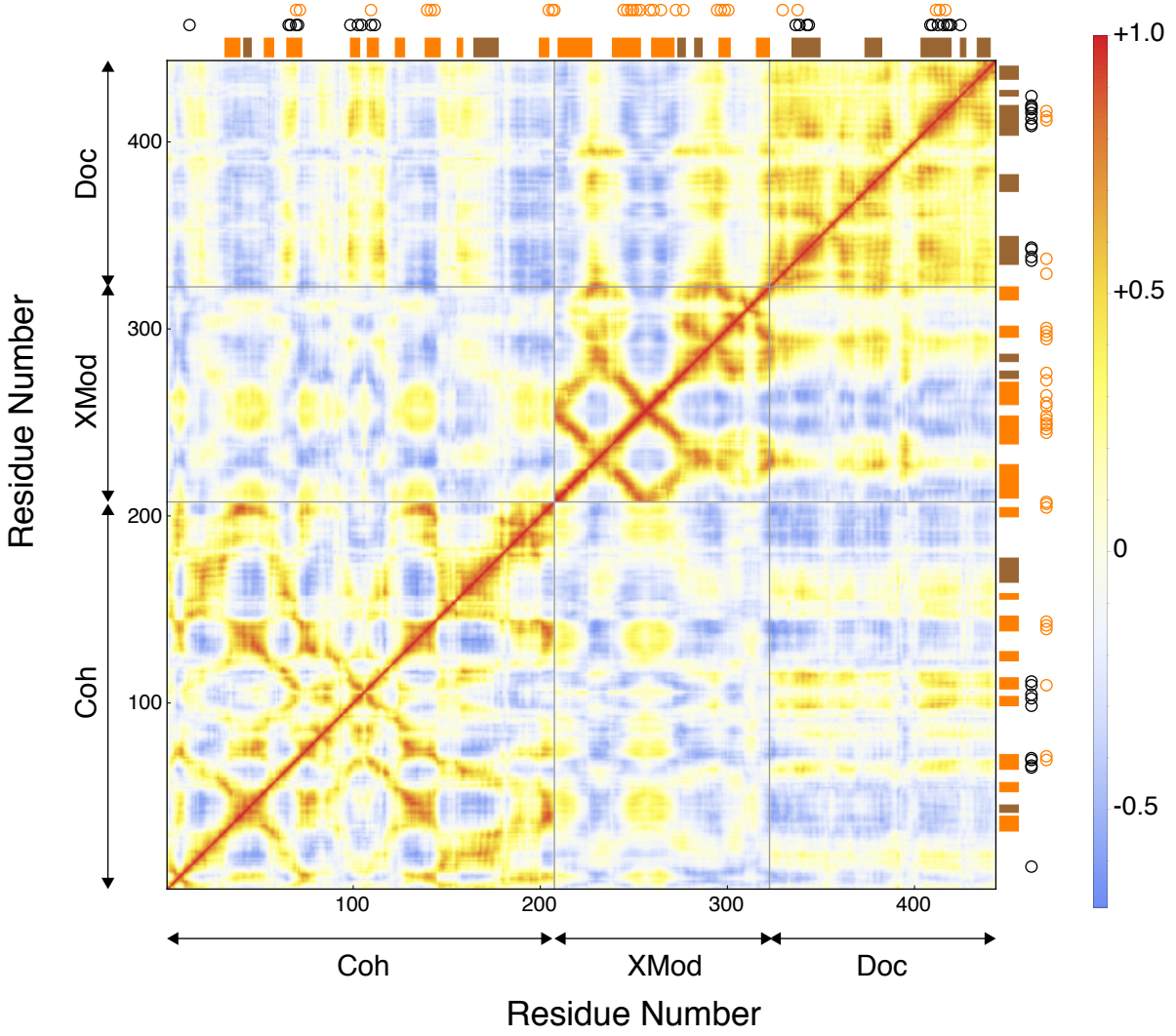


Figure 4.8: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the Xmod-Doc:Coh complex loaded with force in the native pulling geometry. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from Coh and Doc binding interfaces and orange circles represent residues on the force propagation path.

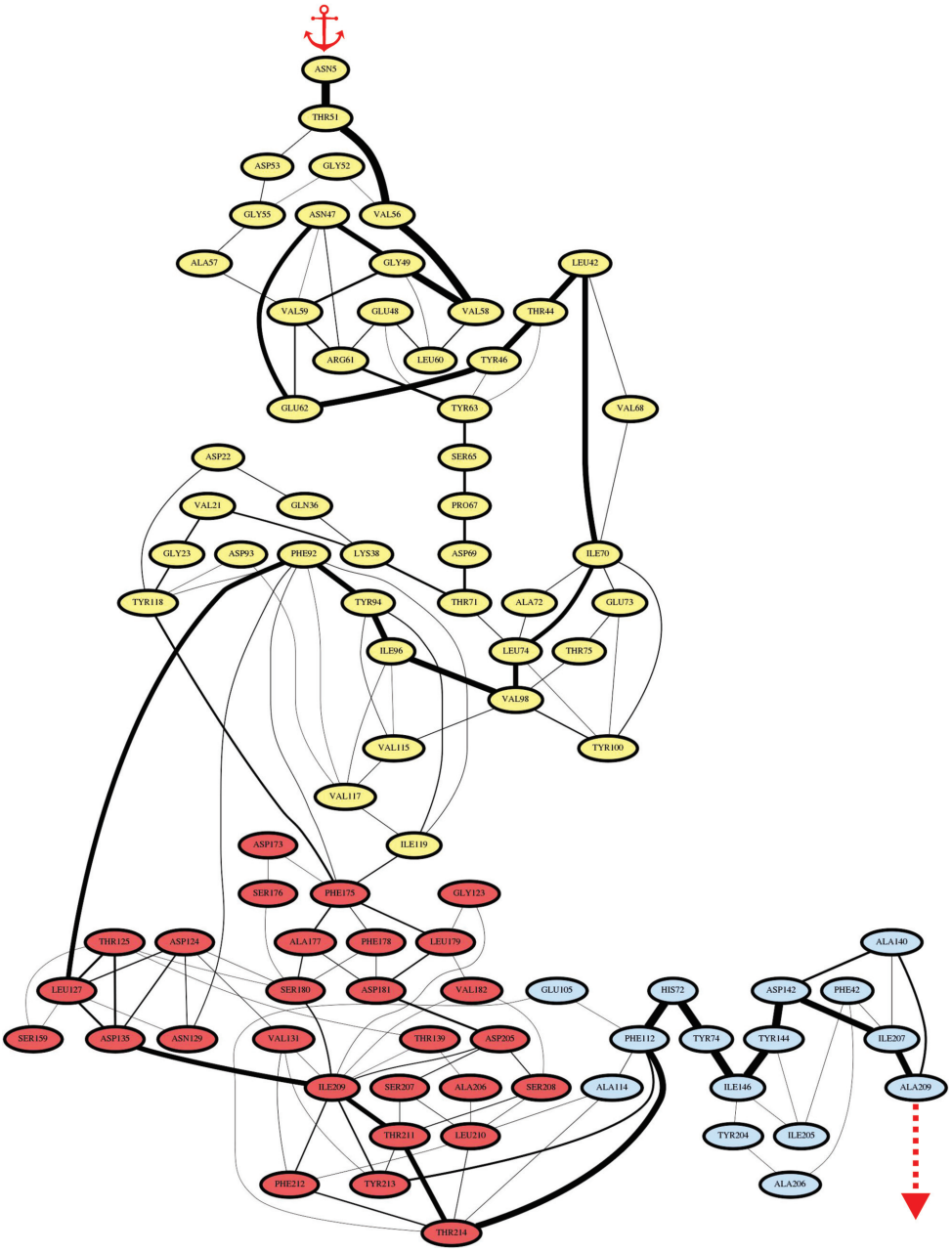


Figure 4.9: Force propagation pathway through the loaded XMod-Doc:Coh complex in the native pulling geometry (N-terminal pulling of Xmod-Doc, C-terminal pulling of Coh) obtained from dynamical network analysis. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

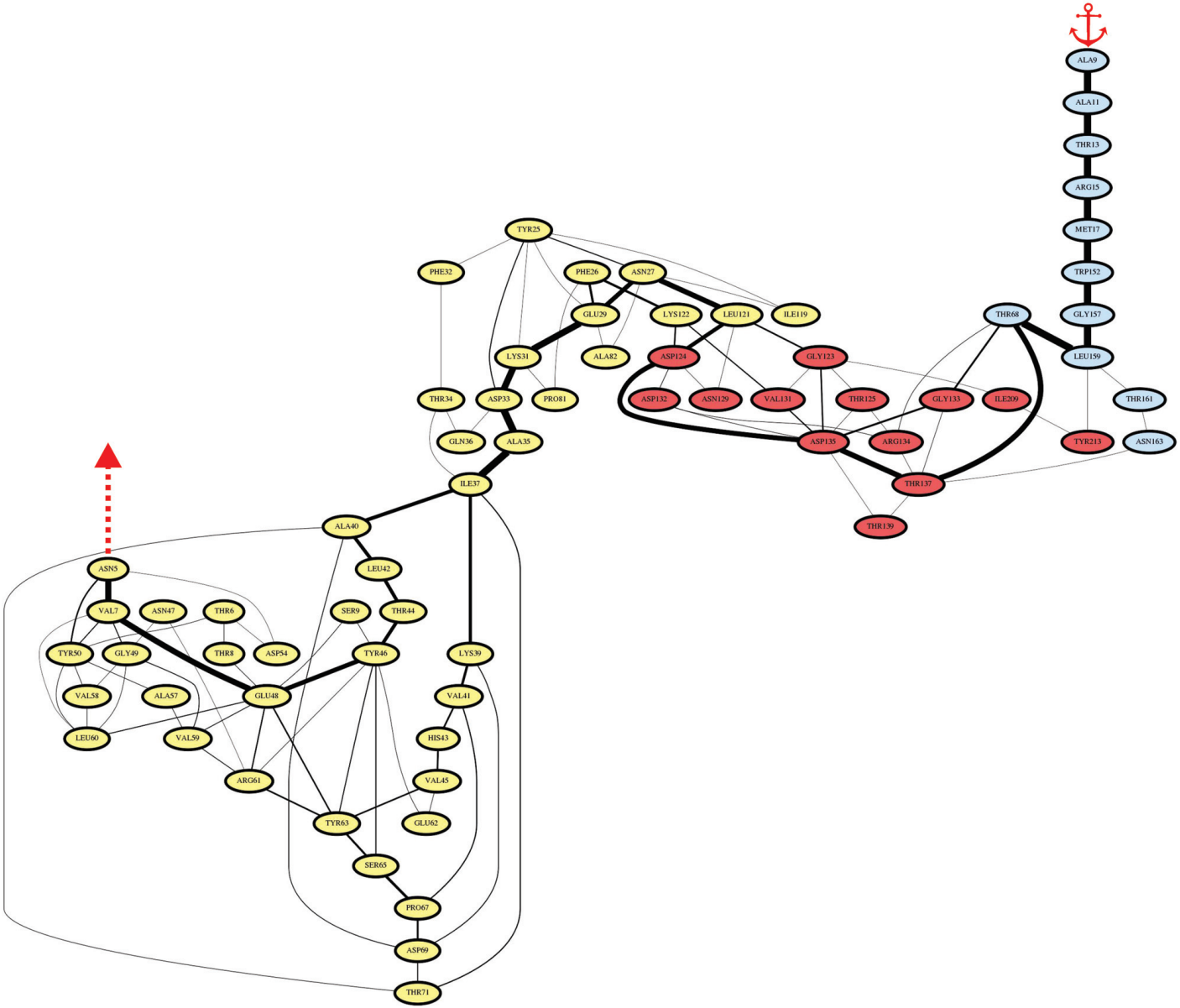


Figure 4.10: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing high-force unbinding characteristics and no C-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

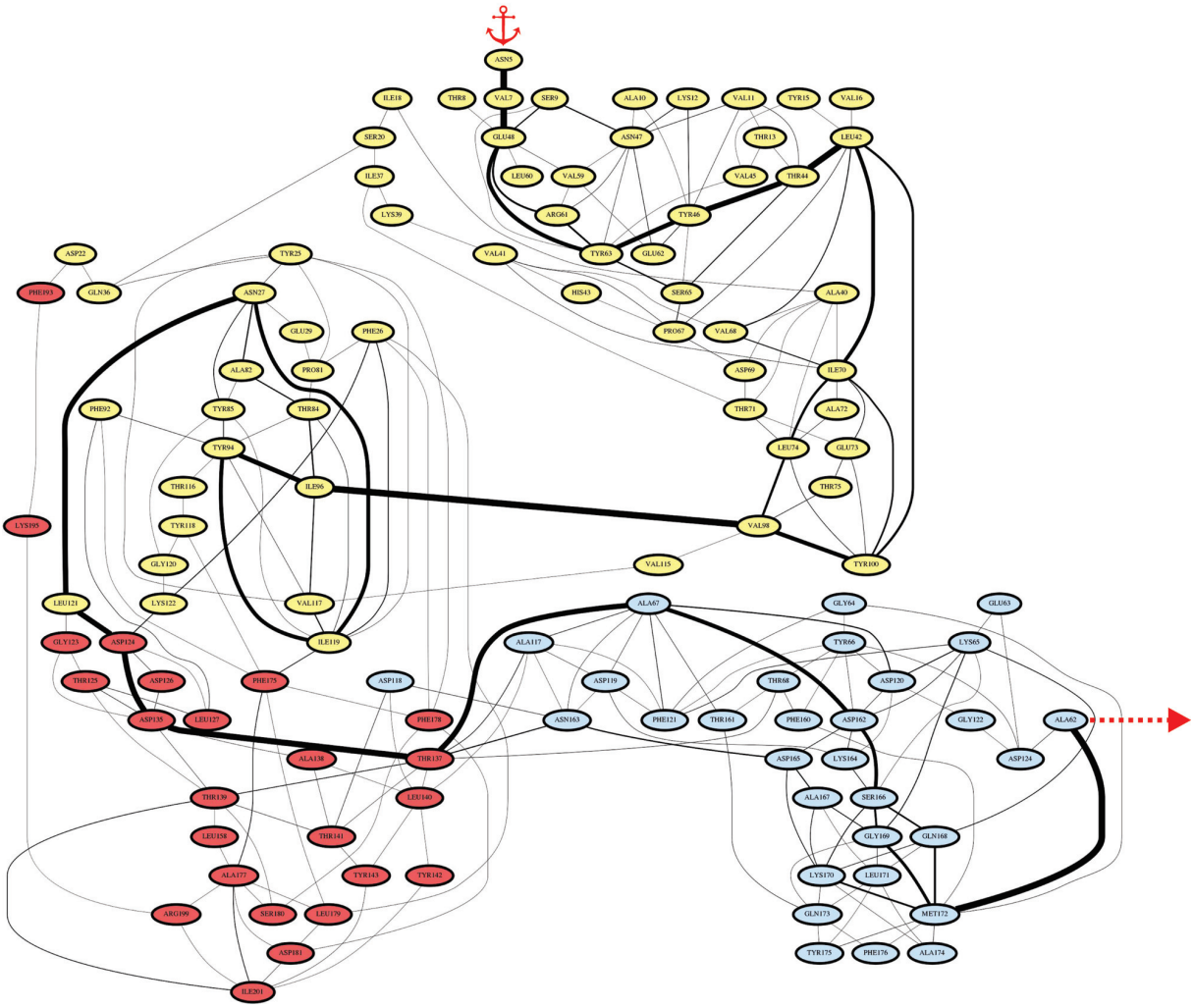


Figure 4.11: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing low-force unbinding characteristics and partial N-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

Figure 4.12: Full unnormalized covariance Matrix M_{ij} for a five atom system from which the full normalized covariance matrix is calculated according to Eq. (4.8). On- and off-diagonal elements from one of the atomic submatrices are highlighted in yellow and blue, respectively.

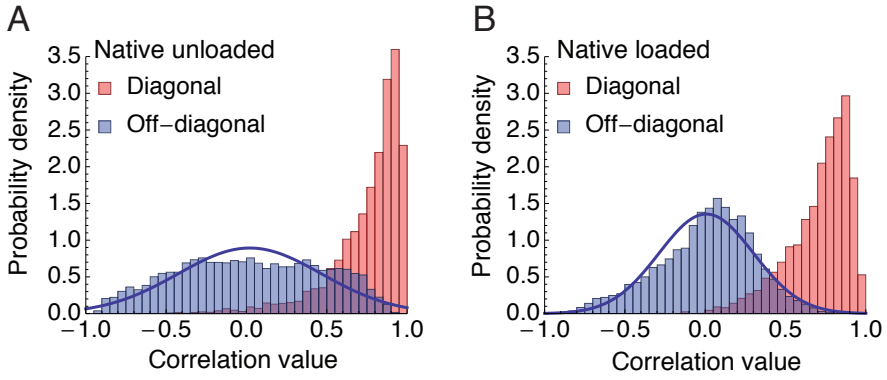


Figure 4.13: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A**, the native unloaded, and **B** the native loaded, scenario.

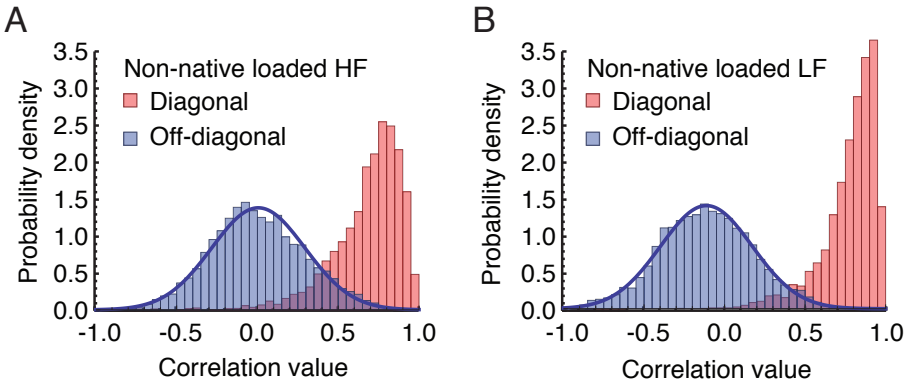


Figure 4.14: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A,** the non-native HF, and **B** the non-native LF, scenario.

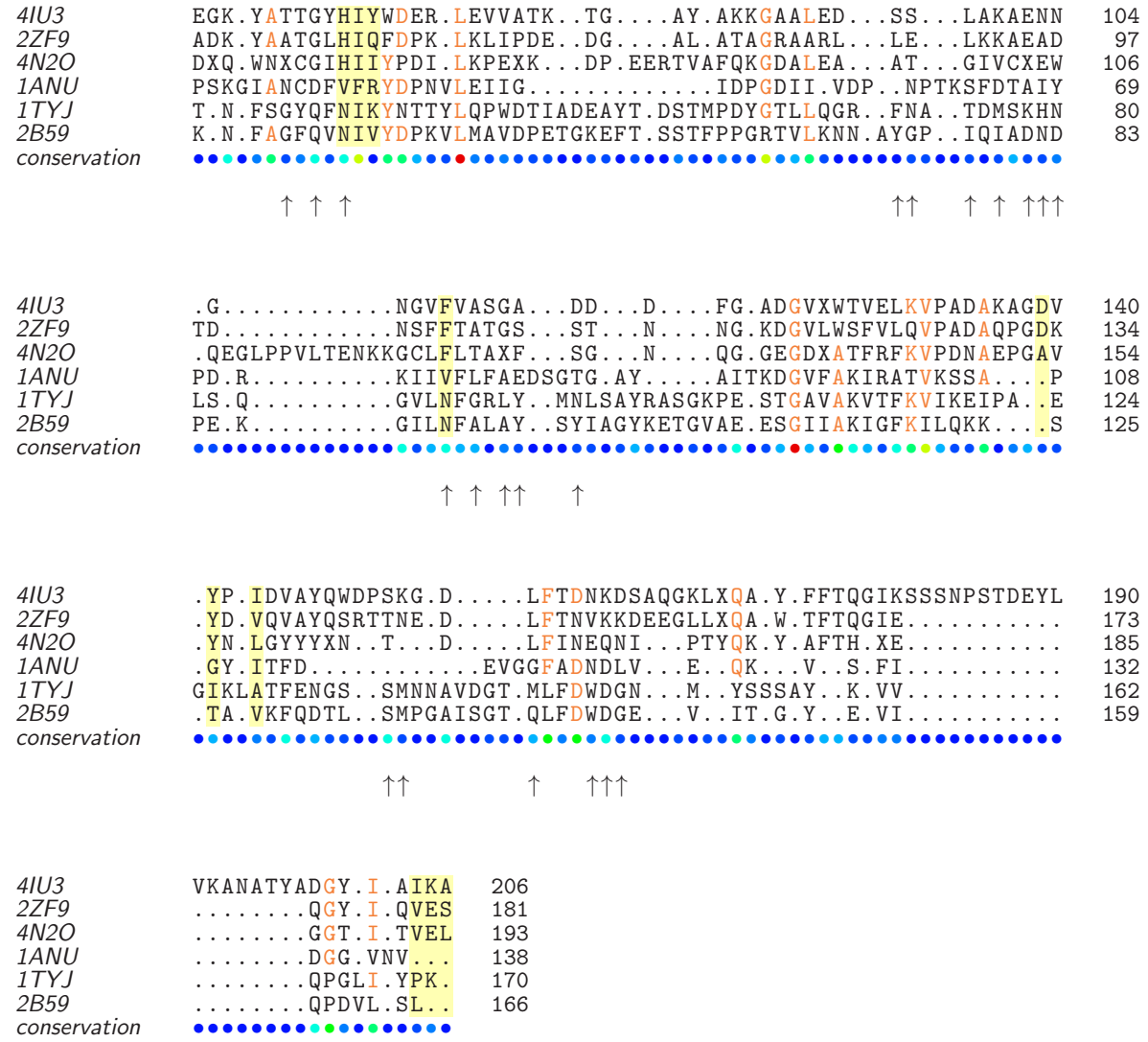


Figure 4.15: Structure-aligned sequences of six crystallized cohesins. Residues on the force propagation path are highlighted in yellow. Arrows indicate binding residues. Residue conservation is color-coded from blue - lack of conservation, to red - residue fully conserved. Crystal structures used: 4IU3 ScaE *Rf* FD-1, 2ZF9 ScaE *Rf* strain 17, 4N2O CohG *Rf* FD-1, 1ANU CohC2 CipC *Ct*, 1TYJ CohA11 ScaA *Bc*, 2B59 SdbA *Ct*.

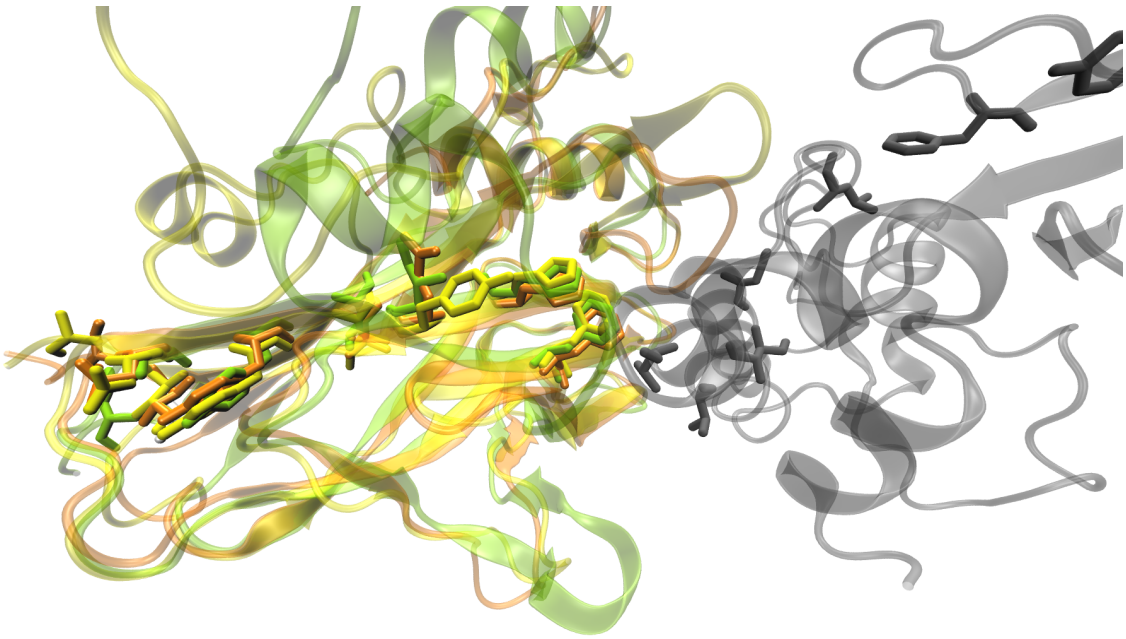


Figure 4.16: Structure and sequence conservation of the force propagation pathway residues in Coh. CohE from the ScaE cell anchoring protein, *Rf* FD-1 used in this work (PDB 4IU3) is highlighted in green. Highly homologous structures of CohE from *Rf* strain 17 (PDB 2ZF9) and Coh G from *Rf* FD-1 (PDB 4A2O) are colored in orange and yellow, respectively. Residues lying in the force propagation path are shown as sticks. XDoc from the CttA *Rf* FD-1 scaffold used in this work is shown in gray.

Chapter 5

Biasing effects of receptor ligand complexes on protein unfolding statistics

Summary

Protein receptor ligand pairs are increasingly used as specific molecular handles in single molecule protein unfolding experiments. Further, known marker domains, also referred to as fingerprints, provide unique unfolding signatures to identify specific single molecule interactions, when receptor ligand pairs themselves are investigated. We show here that in cases where there is an overlap between the probability distributions associated with fingerprint domain unfolding and receptor-ligand dissociation, the experimentally measured force distributions are mutually biased. This biasing effect masks the true parameters of the underlying free energy landscape. To address this, we present a model-free theoretical framework that corrects for the biasing effect caused by such overlapping distributions.

5.1 Introduction

Mechanical forces play an important role in many biological systems. Refolding of individual titin domains is believed to assist in muscle contraction [218], stretching forces expose cryptic binding sites involved in focal adhesions [219], and mechanically stable receptor-ligand pairs govern the assembly of large extracellular machineries and adhesion of bacterial cells to their cellulosic carbon sources [5, 73]. Single molecule pulling experiments with atomic force microscopes [7], optical tweezers [220], or magnetic tweezers [221] have become widely used techniques to study such phenomena at the single molecule level.

Due to the stochastic nature of domain unfolding, typical AFM experiments record many thousands of data traces to obtain statistically meaningful results from single molecule pulling experiments. To unambiguously identify the unfolding signals of a given protein domain of interest, or the dissociation of a receptor-ligand system under external load, the resulting

This Chapter was published by Schoeler *et al.* [80] in *Physical Review E* and adapted with permission from the American Physical Society.

datasets need to be filtered and specific, single molecule interactions must be identified.

To accommodate this need, the community has developed an elegant strategy to achieve high yields of specific curves: In this approach, protein domains of interest are fused to a receptor complex that serves as a specific handle in pulling experiments. This improves curve yields and data fidelity by providing a specific molecular interaction handle to 'grab' the protein of interest. Thereby, the unfolding of individual domains and the dissociation of a receptor ligand complex can be studied in a single experiment [5, 69, 70, 73, 78]. Also, using a known protein domain in the fusion construct provides a unique unfolding pattern that can be used to identify specific traces, when receptor ligand unbinding itself is studied. These domains are then referred to as fingerprint domains [7].

In order for a curve to be unambiguously identified as constituting specific signal, it needs to exhibit unfolding of all included fingerprint domains plus a specific receptor ligand dissociation event. Throughout this letter we will refer to domains fused to a receptor ligand complex as fingerprint domains for both scenarios, namely protein unfolding studies using receptor ligand complexes as specific handles, as well as unbinding studies of receptor-ligand complexes of interest, which use fingerprint domains for assistance in data filtering. We discuss the statistical effects that arise when the respective force distributions for fingerprint domain unfolding and receptor ligand complex rupture exhibit a finite overlap. We quantitatively show how the statistical overlap between these two distributions affects the experimentally observable unfolding and rupture force distributions. We provide a framework for extracting kinetic and energetic information from the experimentally observed distributions that are corrected for the biasing effects arising from such overlaps in a model free fashion.

5.2 Theoretical framework

The standard theoretical framework treats protein unfolding or bond dissociation as thermally driven escape over a free energy barrier that is modulated by an external force F [118–121]. This description leads to a general expression for the distribution of unfolding or rupture forces in a pulling experiment:

$$p\left(F\right) = \frac{k\left(F\right)}{\dot{F}\left(F\right)} \exp\left(-\int_0^F \mathrm{d}F' \, \frac{k\left(F'\right)}{\dot{F}\left(F'\right)}\right) \tag{5.1}$$

where $k\left(F\right)$ is the force dependent off rate of the system, and $\dot{F}\left(F\right)$ is the force loading rate. In the simplest picture[119, 120], the force dependent off rate is given by[118]:

$$k\left(F\right) = k_0 \exp\left(\frac{F\Delta x}{k_\mathrm{B}T}\right) \tag{5.2}$$

where k_0 and Δx are the zero force off rate and distance to the free energy barrier, respectively. For a constant force loading rate $\dot{F}$, and an off rate from Eq. (5.2), the integral in Eq. (5.1) can be solved analytically (Fig. 5.1A). Dudko *et al.* [123] have used the Kramers theory [124] to obtain a more sophisticated expression for the force dependent off rate, which accounts for the specific shape of the free energy landscape. This approach also provides an analytical solution to Eq. (5.1) for a constant loading rate and includes the height of the free energy barrier ΔG as

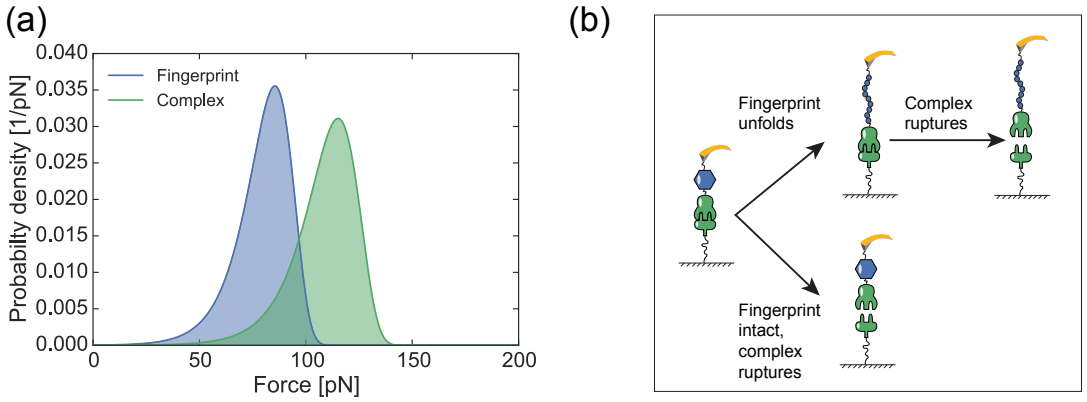


Figure 5.1: (A) Overlapping distributions of fingerprint unfolding (blue) and complex rupture (green) for a constant loading rate $\dot{F} = 200\,\mathrm{pN\,s^{-1}}$ with fingerprint $\Delta x = 0.4\,\mathrm{nm}$, and $k_0 = 0.005\,\mathrm{s^{-1}}$ and complex parameters $\Delta x = 0.35\,\mathrm{nm}$, and $k_0 = 0.001\,\mathrm{s^{-1}}$ obtained by integrating Eq. (5.1) using Eq. (5.2). (B) Schematic of possible outcomes for the situation in (A). Analyzable data show fingerprint unfolding followed by complex rupture (upper path). Due to the overlapping distributions for unfolding and rupture, complex rupture with intact fingerprint is also possible (lower path).

an additional parameter. Over the years, more theoretical models describing the distributions for domain unfolding and receptor ligand dissociation have been developed [117, 122, 222–225] that can be applied to experimentally measured force distributions.

Since a constant force loading rate is required to obtain an analytical solution for the distribution of forces in a pulling experiment, force ramp measurements [126, 226], where the external force is ramped linearly, are an elegant technique to study protein unfolding and receptor ligand dissociation. In an experiment where a receptor ligand system is used to probe the unfolding behavior of a protein fingerprint domain of interest, care has to be taken when analyzing the resulting unfolding or rupture force distributions. If the probability distributions for fingerprint domain unfolding and complex rupture are disjunct, *i.e.* the complex ruptures at forces well above those required for fingerprint unfolding, the measured distributions are unbiased and can be readily analyzed using a desired form of Eq. (5.1). If those distributions have a substantial overlap (Fig. 5.1A), however, biasing effects have to be taken into account.

To pass the data analysis filter a given curve is required to exhibit both fingerprint unfolding and complex rupture (Fig. 5.1B, upper path), *i.e.* the complex must not rupture prior to fingerprint unfolding (Fig. 5.1B, lower path). Consequently, the resulting distribution of fingerprint unfolding forces will be shifted downwards towards lower forces. By the same logic, experimentally observed distributions for receptor-ligand complex rupture forces will be truncated at the lower end and shifted upwards toward higher forces in a constant loading rate experiment. This biasing effect has been used qualitatively by Jobst *et al.* [69] to unambiguously identify a redundant dual binding mode in a receptor ligand complex.

Here we treat this biasing effect quantitatively and calculate these effects independent of the model used in Eq. (5.1). For overlapping distributions of fingerprint, $p_\mathrm{f}(F)$, and receptor ligand complex, $p_\mathrm{c}(F)$, the biased distribution of the fingerprint, $p_\mathrm{f}^\star(F)$, can be calculated by multiplying the original distribution with the cumulative probability for the bond to rupture at

higher forces and renormalizing:

$$p_{\mathrm{f}}^{\star}(F) = \frac{p_{\mathrm{f}}(F) \int_{F}^{\infty} \mathrm{d}F'\, p_{\mathrm{c}}(F')}{\eta} \tag{5.3}$$

where η is a normalization constant. Since $p_{\mathrm{c}}(f)$ is normalized, Eq. (5.3) can be rewritten:

$$p_{\mathrm{f}}^{\star}(F) = \frac{p_{\mathrm{f}}(F)\left(1 - \int_{0}^{F} \mathrm{d}F'\, p_{\mathrm{c}}(F')\right)}{\eta} \tag{5.4}$$

$$= \frac{p_{\mathrm{f}}(F)\,(1 - P_{\mathrm{c}}(F))}{\eta} \tag{5.5}$$

where $P_{\mathrm{c}}(F)$ is the cumulative distribution function of the complex rupture forces. The normalization constant η can be calculated by integrating over the numerator in Eq. (5.5):

$$\eta = 1 - \int_{0}^{\infty} \mathrm{d}F\, p_{\mathrm{f}}(F)\, P_{\mathrm{c}}(F) \tag{5.6}$$

Intuitively, the biased fingerprint distribution is normalized by the ratio of curves that exhibit fingerprint unfolding vs. all rupture events. The above calculations apply for both constant loading rate and constant speed measurements. By the same logic, the biased distribution of observed complex ruptures for a constant loading rate can be calculated as:

$$p_{\mathrm{c}}^{\star}(F) = \frac{p_{\mathrm{c}}(F)\, P_{\mathrm{f}}(F)}{\eta} \tag{5.7}$$

Both biased distributions for fingerprint unfolding, $p_{\mathrm{f}}^{\star}(F)$, and complex ruptures, $p_{\mathrm{c}}^{\star}(F)$, are normalized by the same yield parameter η since both distributions are extracted from the same curves in a given data set. For a mathematical proof, see Appendix 5.6.1. We note that the biasing effect on complex rupture in the constant speed case is more difficult to quantify. Since the additional contour length released upon fingerprint unfolding is not immediately compensated for by a feedback on the force signal, there is a pronounced drop in force associated with fingerprint unfolding giving rise to the characteristic sawtooth pattern in force extension traces. Ignoring the finite relaxation time of the harmonic pulling device, the force will drop from $F_1 = F(x, L)$ to $F_2 = F(x, L + \Delta L)$ where the former describes the (non-linear) elastic behavior of the pulling device and potential linker molecules, and ΔL is the additional contour length released upon fingerprint unfolding. Whether or not the observed complex distribution is biased, depends on the value of $p_{\mathrm{c}}(F_2)$. For $p_{\mathrm{c}}(F_2) \approx 0$, no biasing will occur, whereas $p_{\mathrm{c}}(F_2) > 0$ will cause a biasing effect. In practice, ΔL is usually large enough to ensure complex distributions are unbiased in constant speed experiments, leaving a substantial bias only on the observed fingerprint distribution. A strategy to implement our correction procedure for the constant speed protocol is proposed in Appendix 5.6.2.

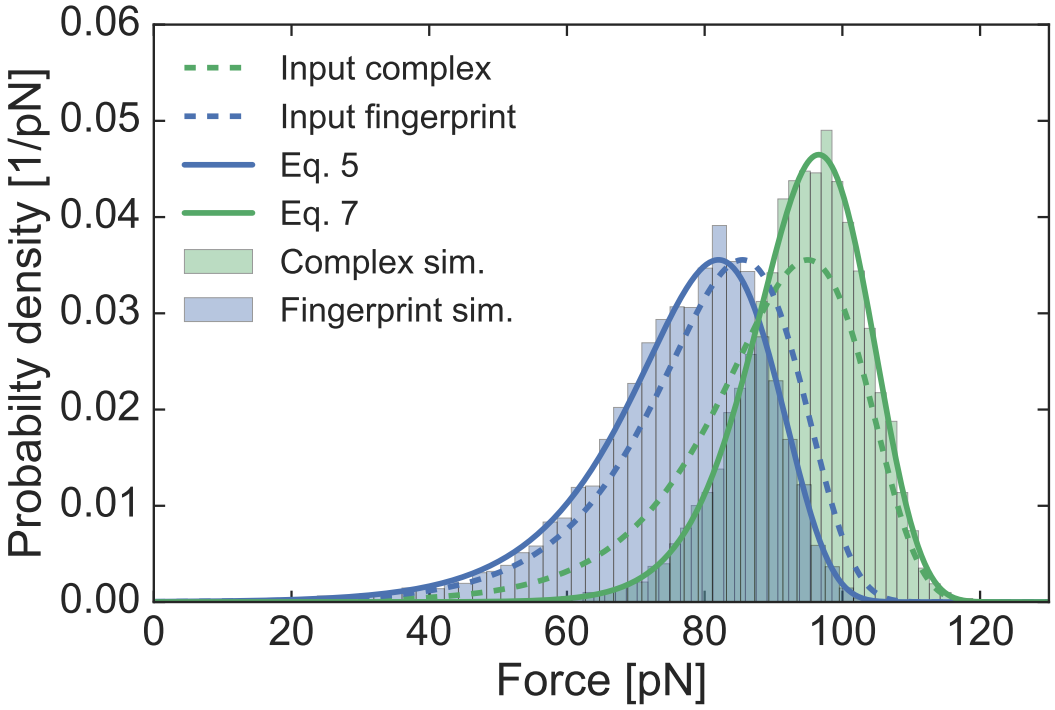


Figure 5.2: Force ramp simulation with $\dot{F} = 200\,\mathrm{pN\,s^{-1}}$, complex $\Delta x_\mathrm{c} = 0.4\,\mathrm{nm}$, $k_{0,\mathrm{c}} = 0.002\,\mathrm{s^{-1}}$ and fingerprint $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$, $k_{0,\mathrm{f}} = 0.005\,\mathrm{s^{-1}}$ ($\eta = 0.71$). Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines are the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5.5) and (5.7) are shown as solid blue and green lines, respectively.

5.3 Monte Carlo simulations

To validate our results, we used a Monte Carlo approach to simulate fingerprint domain unfolding in combination with receptor ligand dissociation in a constant loading rate protocol. Our simulation routine is similar to the approach described in ref. [227] and uses the phenomenological model due to its analytical tractability. Briefly, we integrate Eq. (5.2) over a time step Δt, where $F = F(t) = \dot{F}t$, to obtain probabilities p_u and p_r for fingerprint unfolding and complex rupture, respectively. These probabilities are compared to independent random numbers n_u and n_r between zero and unity. If $p_\mathrm{u} < n_\mathrm{u}$ and $p_\mathrm{r} < n_\mathrm{r}$, fingerprint and complex remain intact and the next iteration is started. If $p_u > n_u$ and $p_r < n_r$ the corresponding force is noted as the fingerprint unfolding force and the next iteration is started with only the complex intact. The simulation then continues until $p_\mathrm{r} > n_\mathrm{r}$ and the corresponding force is noted as the complex rupture force. If $p_\mathrm{u} < n_\mathrm{u}$ and $p_\mathrm{r} > n_\mathrm{r}$ or $p_\mathrm{u} > n_\mathrm{u}$ and $p_\mathrm{r} > n_\mathrm{r}$, the complex ruptured prior to fingerprint unfolding or at the same time, and an experimental curve would be unanalyzable and filtered out during the data analysis procedure.

Results from a Monte Carlo simulation at constant loading rate $\dot{F} = 200\,\mathrm{pN\,s^{-1}}$ with overlapping distributions for fingerprint unfolding and complex rupture are shown in Fig. 5.2. $N = 10000$ traces were simulated and the observed fingerprint domain unfolding forces and receptor-ligand complex rupture forces were histogrammed. We only analyzed curves that showed both fingerprint unfolding and complex rupture to mimic experimental conditions. As expected, both fingerprint unfolding and complex rupture distributions (blue and green histograms in Fig. 5.2) are shifted from the unbiased input distributions (shown as blue and green dashed lines in Fig. 5.2). The biased results are well described by our theoretical predictions, which are shown as green and blue and green solid lines in Fig. 5.2. To illustrate the potential errors that can occur from not accounting for the fingerprint biasing effect, we used the uncorrected distribution (Eq. (5.1)) to fit the biased fingerprint domain unfolding and complex rupture histograms and compared the resulting fit parameters Δx_fit and $k_{0,\mathrm{fit}}$ with the unbiased input parameters (Table 5.1). We found that for fingerprint domain unfolding, Δx is hardly affected, while k_0 is overestimated by 30 %. For complex rupture Δx is overestimated by 29 % due to the smaller width of the biased distribution, while k_0 is underestimated by over an order of magnitude. If the unbiased parameters for the complex distribution are known from a control experiment, our predicted biased distributions can be used to fit the experimental data to obtain unbiased values for the fit parameters pertaining to the fingerprint or vice versa. Alternatively, a global fitting procedure can be applied to both biased distributions for constant loading rate data to obtain unbiased fit parameters without prior knowledge of either distribution. Using this approach, we obtained global fit parameters that were within 6 % of the input parameters

Table 5.1: Input vs. fit parameters of simulation shown in Fig. 5.2. For these parameters, the yield parameter equals $\eta = 0.71$. Initially, the simulated distributions were fit with the uncorrected distributions ($k_{0,\mathrm{fit}}$ and Δx_fit). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions (Eqs. 5.5 and 5.7) in a global fitting procedure to obtain the corrected parameters $k_{0,\mathrm{global}}$ and Δx_global.

	$k_{0,\mathrm{in}}\left[\mathrm{s^{-1}}\right]$	$k_{0,\mathrm{fit}}\left[\mathrm{s^{-1}}\right]$	$k_{0,\mathrm{global}}\left[\mathrm{s^{-1}}\right]$	Δx_in [nm]	Δx_fit [nm]	Δx_global [nm]
Fingerprint	5.0×10^{-3}	$(6.6 \pm 0.2) \times 10^{-3}$	$(4.7 \pm 0.4) \times 10^{-3}$	0.40	0.40 ± 0.01	0.40 ± 0.01
Complex	2.0×10^{-3}	$(0.1 \pm 0.1) \times 10^{-3}$	$(1.9 \pm 0.2) \times 10^{-3}$	0.40	0.53 ± 0.01	0.40 ± 0.01

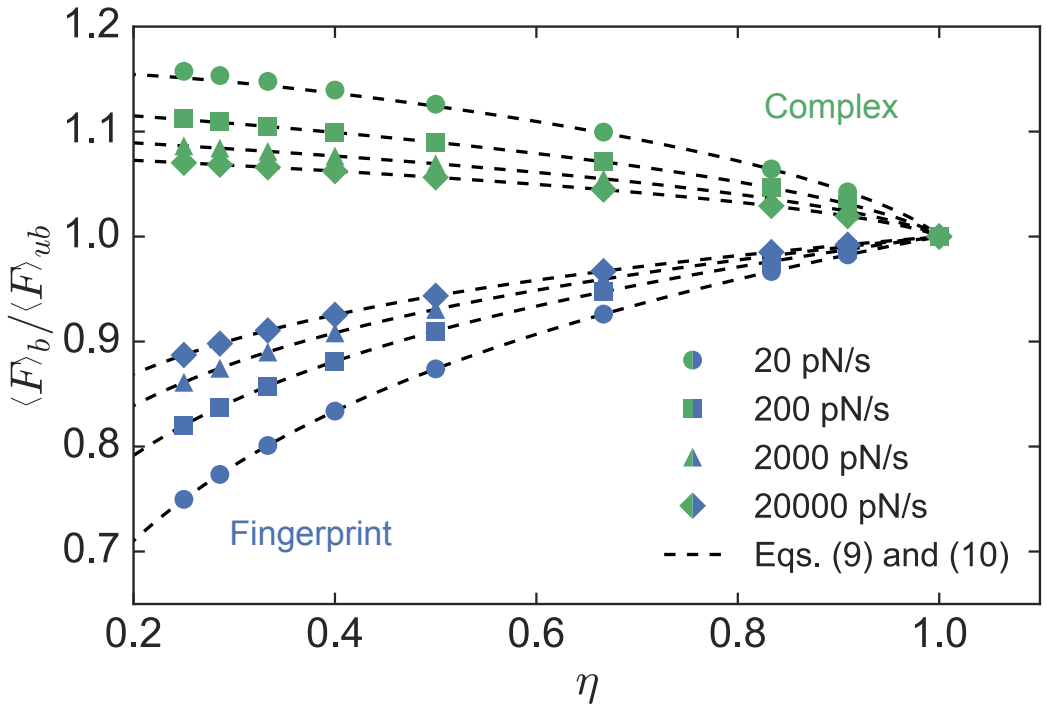


Figure 5.3: Biasing of fingerprint unfolding and complex rupture forces as a function of the theoretical yield of curves exhibiting fingerprint unfolding η. Data points show relative mean biased unfolding force $\langle F\rangle_\mathrm{b}$ / $\langle F\rangle_\mathrm{ub}$, fingerprint and complex data are shown in blue and green, respectively. For the fingerprint data, parameters were held constant at $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$ and $k_{0,\mathrm{f}} = 0.005\,\mathrm{s}^{-1}$ while the complex distribution was shifted by maintaining $\Delta x_\mathrm{c} = 0.4\,\mathrm{nm}$ and varying the off rate. For the complex data, $\Delta x_\mathrm{c} = 0.4\,\mathrm{nm}$ and $k_{0,\mathrm{c}} = 0.005\,\mathrm{s}^{-1}$ were held constant and the fingerprint distribution was shifted by maintaining $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$ and varying the off rate. Dashed lines are predictions based on Eq. (5.9) and (5.10).

(*cf.*, Table5.1).

5.4 Special case: equal potential widths

In Fig. 5.3 we quantify the magnitude of the biasing effect by numerically calculating the relative mean biased unfolding force of the fingerprint (blue symbols) and complex (green symbols) $\langle F\rangle_\mathrm{b} / \langle F\rangle_\mathrm{ub} = \int \mathrm{d}F\, F p^*_{\mathrm{f/c}}(F) / \int \mathrm{d}F\, F p_{\mathrm{f/c}}(F)$ as a function of the theoretical ratio of curves exhibiting fingerprint unfolding prior to receptor-ligand complex rupture, η (Eq. (5.6). For analytical tractability we chose the special case $\Delta x_\mathrm{c} = \Delta x_\mathrm{f}$. With this simplification we find for the fingerprint:

$$\langle F\rangle_\mathrm{b} = \frac{k_\mathrm{B}T}{\Delta x_\mathrm{f}} e^{\frac{k_\mathrm{B}Tk_{0,\mathrm{f}}}{\Delta x_\mathrm{f} r\eta}} E_1\left(\frac{k_\mathrm{B}Tk_{0,\mathrm{f}}}{\Delta x_\mathrm{f} r\eta}\right) \tag{5.8}$$

where $E_1(x)$ is the exponential integral. Using $e^x E_1(x) \cong \ln\left(1+\frac{e^{-\gamma}}{x}\right)$ we can produce an analytical approximation for the relative mean biased unfolding force for the aforementioned

special case:

$$\frac{\langle F\rangle_{\mathrm{b}}}{\langle F\rangle_{\mathrm{ub}}} \cong \frac{\ln\left(1+\frac{r\Delta x_{\mathrm{f}}}{k_{0,\mathrm{f}}k_{\mathrm{B}}T}e^{-\gamma}\eta\right)}{\ln\left(1+\frac{r\Delta x_{\mathrm{f}}}{k_{0,\mathrm{f}}k_{\mathrm{B}}T}e^{-\gamma}\right)} \tag{5.9}$$

where $\gamma = 0.577\ldots$ is the Euler-Mascheroni constant. The analogous result for the complex reads:

$$\frac{\langle F\rangle_{\mathrm{b}}}{\langle F\rangle_{\mathrm{ub}}} \cong \frac{1}{\eta} - \frac{\ln\left(1+\frac{r\Delta x_{\mathrm{c}}}{k_{0,\mathrm{c}}k_{\mathrm{B}}T}e^{-\gamma}\,(1-\eta)\right)}{\ln\left(1+\frac{r\Delta x_{\mathrm{c}}}{k_{0,\mathrm{c}}k_{\mathrm{B}}T}e^{-\gamma}\right)}\frac{1-\eta}{\eta} \tag{5.10}$$

For this special case Eq. (5.6) can be evaluated analytically yielding $\eta = \left(1+\frac{k_{0,\mathrm{c}}}{k_{0,\mathrm{f}}}\right)^{-1}$. Eqs. (5.9) and (5.10) (dashed lines in Fig. 5.3) agree very well with our numerical results over the loading rate regime covered. Fig. 5.3 clearly shows that the biasing effect is more pronounced for small loading rates, consistent with our theoretical predictions based on Eqs. (5.9) and (5.10). In cases where data cannot be corrected for a potential biasing effect, *e.g.* due to low experimental yields, the magnitude of the biasing effect can be minimized by increasing the loading rate in a pulling experiment.

5.5 Conclusion

Our calculations provide a framework for analyzing single molecule force spectroscopy data where receptor ligand systems are used as specific handles to study a fingerprint domain of interest or vice versa. In such experiments, it is many times the case that the distributions of fingerprint domain unfolding and complex rupture have a significant overlap (a few exemplary cases can be found in refs. [3, 70, 73, 78]). In this case biasing effects will occur and should be considered in the analysis procedure. Our findings can be applied to both constant speed and force ramp (constant loading rate) experimental protocols, however, it should be noted that the biasing effect on complex unbinding is more complicated in the constant speed protocol, due to the drop in force upon fingerprint unfolding dependent on the length of the unfolded domain. Since the biasing effects are solely due to the statistical nature of domain unfolding or complex unbinding, our results, specifically Eqs. (5.5) and (5.7), are valid regardless of the specific model used in Eq. (5.1).

5.6 Appendix

5.6.1 Identity of yield parameter η in Eqs. (5.5) and (5.7)

As we state in the main text, the normalization parameter η is equal to the ratio of curves that exhibit fingerprint unfolding vs. all rupture events. In other words a fraction $1-\eta$ of all curves will be "missed" in an actual experiment, since they do not exhibit fingerprint unfolding and would hence be discarded during data analysis. Consequently, both fingerprint and complex distributions (which are obtained from the same curves) need to be normalized by the same η to obtain probability distributions normalized to unity. Mathematically, one needs to show

that:

$$1 - \int_0^\infty \mathrm{d}F\, p_\mathrm{f}(F)\, P_\mathrm{c}(F) = \eta = \int_0^\infty \mathrm{d}F\, p_\mathrm{c}(F)\, P_\mathrm{f}(F) \tag{5.11}$$

We use integration by parts and $\int_0^F \mathrm{d}F'\, p(F') = P(F)$ to evaluate the rhs of Eq. (5.11):

$$\mathrm{rhs} = P_\mathrm{c}(F)\, P_\mathrm{f}(F)\Big|_0^\infty - \int_0^\infty \mathrm{d}F\, P_\mathrm{c}(F)\, p_\mathrm{f}(F) \tag{5.12}$$

Since $P(0) = 0$ and $P(\infty) = 1$ this is equal to the lhs of Eq. (5.11).

5.6.2 Constant speed

As pointed out in the main text, the correction procedure for distributions obtained from constant speed measurements is more involved. In this Appendix we discuss a strategy for extracting unbiased parameters from fingerprint distributions for this pulling protocol. Due to the non-linear elasticity of linker molecules (e.g. P(oly)E(thylene)G(lycol), spacers or unfolded protein backbone), the force loading rate $\dot{F} = \dot{F}(F)$ becomes a function of the force and the integral in Eq. (5.1) can no longer be evaluated analytically. In a standard pulling experiment, a harmonic pulling device (*e.g.* AFM cantilever or optically trapped bead) is connected to aforementioned linker molecules. By applying a force balance it can be shown that the force loading rate is given by[121]:

$$\frac{v}{\dot{F}(F)} = \frac{1}{k_\mathrm{h}} + \frac{\partial x(F)}{\partial F} \tag{5.13}$$

where v is the pulling speed, k_h is the spring constant of the harmonic pulling device, and $x(F)$ is the force dependent extension of the linker. The biased distribution for fingerprint unfolding $p_\mathrm{f}^\star(F)$ can be computed by numerically solving the integrals occurring in Eqs. (5.5) and (5.6), using a model for the elastic response of the linker molecules such as the worm-like-chain model, freely rotating model or the composite model proposed by Livadaru *et al.* [113] in Eq. (5.13). The choice of model should be made based upon the force regime in which fingerprint unfolding and complex rupture are expected and the molecular linkers present in an experimental setup need to be accounted for in these models via their contour length, L, and elasticity, *e.g.* persistence length p. The force dependent loading rate for the worm-like chain model has been derived as Eq. (4) in ref. [121].

5.6.3 Supplemental figures and tables

This Appendix contains contains two figures and one table that support the results of the main text. Fig. 5.4 shows the results of a simulation with identical parameters as in Fig. 5.2, except only $N = 1000$ traces were simulated to mimic a total number of curves more similar to average experimental yields. The extracted fit parameters for this simulation are shown in Table 5.2. We note that albeit an increase in uncertainties, the extracted parameters from our global fitting procedure still reproduce the input parameters much better than those obtained from a fit to the uncorrected distributions. Fig. 5.5 shows the results of Monte Carlo simulations at

different loading rates ranging from $\dot{F} = 20\,\mathrm{pN\,s^{-1}}$ to $\dot{F} = 20\,000\,\mathrm{pN\,s^{-1}}$. Due to the different potential widths $\Delta x_\mathrm{c} = 0.35\,\mathrm{nm}$ and $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$, the yield parameter η varies for the different simulations.

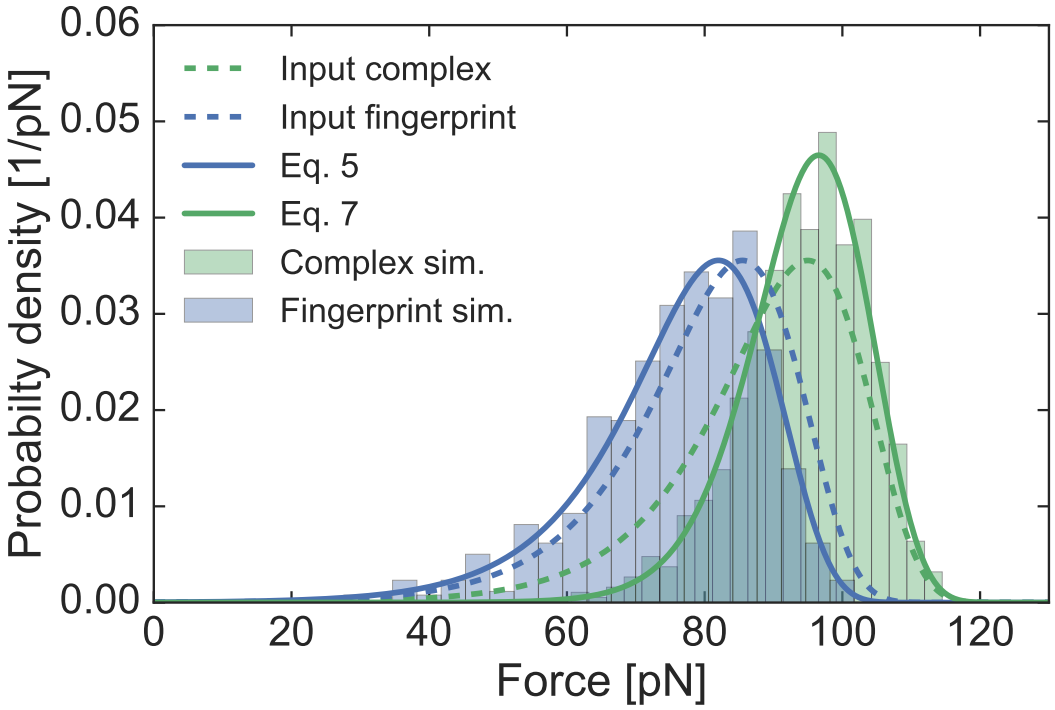


Figure 5.4: Force ramp simulation with $\dot{F} = 200\,\mathrm{pN\,s^{-1}}$, complex $\Delta x_\mathrm{c} = 0.4\,\mathrm{nm}$, $k_{0,\mathrm{c}} = 0.002\,\mathrm{s^{-1}}$ and fingerprint $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$, $k_{0,\mathrm{f}} = 0.005\,\mathrm{s^{-1}}$ ($\eta = 0.71$), and simulated traces $N = 1000$. Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines are the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5.5) and (5.7) are shown as solid blue and green lines, respectively.

Table 5.2: Input vs. fit parameters of simulation shown in Fig. 5.4. For these parameters, the yield parameter equals $\eta = 0.71$. Initially, the simulated distributions were fit with the uncorrected distributions ($k_{0,\mathrm{fit}}$ and Δx_fit). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions (Eqs. 5.5 and 5.7) in a global fitting procedure to obtain the corrected parameters $k_{0,\mathrm{global}}$ and Δx_global.

	$k_{0,\mathrm{in}}\,[\mathrm{s^{-1}}]$	$k_{0,\mathrm{fit}}\,[\mathrm{s^{-1}}]$	$k_{0,\mathrm{global}}\,[\mathrm{s^{-1}}]$	Δx_in [nm]	Δx_fit [nm]	Δx_global [nm]
Fingerprint	5.0×10^{-3}	$(5.9 \pm 1.2) \times 10^{-3}$	$(4.3 \pm 1.5) \times 10^{-3}$	0.40	0.41 ± 0.01	0.41 ± 0.02
Complex	2.0×10^{-3}	$(0.2 \pm 0.1) \times 10^{-3}$	$(2.7 \pm 1.3) \times 10^{-3}$	0.40	0.50 ± 0.02	0.38 ± 0.02

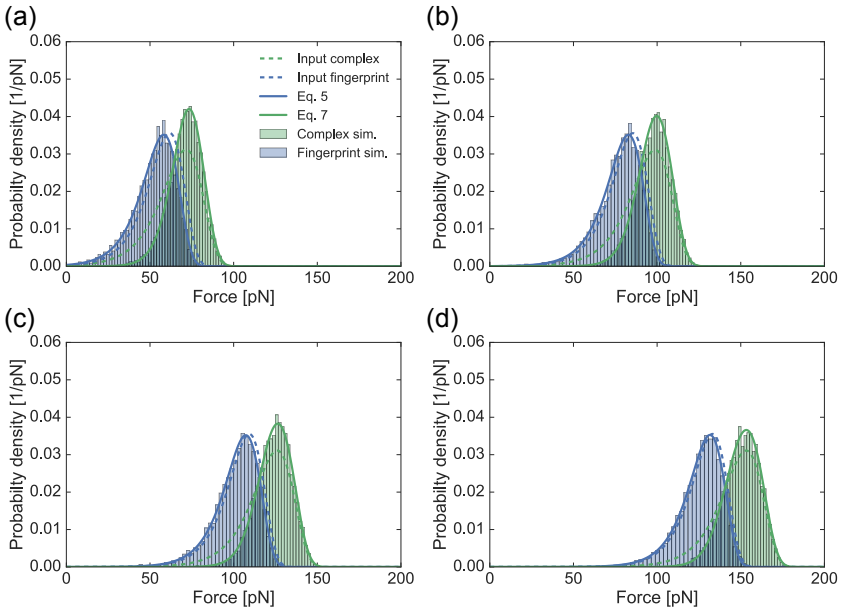


Figure 5.5: Force ramp simulation with complex $\Delta x_\mathrm{c} = 0.35\,\mathrm{nm}$, $k_{0,\mathrm{c}} = 0.004\,\mathrm{s}^{-1}$ and fingerprint $\Delta x_\mathrm{f} = 0.4\,\mathrm{nm}$, $k_{0,\mathrm{f}} = 0.005\,\mathrm{s}^{-1}$, and varying $\dot{F}$. Due to the different potential widths Δx, the yield parameter η changes width the loading rate. Panels (a)-(d) correspond to values of the loading rate and resulting yield parameter of $\dot{F} = 20\,\mathrm{pN\,s}^{-1}$ and $\eta = 0.70$, $\dot{F} = 200\,\mathrm{pN\,s}^{-1}$ and $\eta = 0.75$, $\dot{F} = 2000\,\mathrm{pN\,s}^{-1}$ and $\eta = 0.80$, and $\dot{F} = 20\,000\,\mathrm{pN\,s}^{-1}$ and $\eta = 0.85$, respectively. Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines are the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5.5) and (5.7) are shown as solid blue and green lines, respectively.

Chapter 6
Conclusion and Outlook

The central aim of this work was to shed light on the molecular mechanisms responsible for mechanical stability in a model receptor ligand complex. This molecular system, namely a type III cohesin dockerin complex, is responsible for anchoring a ruminal bacterium to cellulosic substrates. Due to the turbulent environment of the rumen, it is expected to be subjected to mechanical forces in its native function caused by shear flow gradients. When probed with an atomic force microscope, it was found that this system is able to withstand force up to ~ 700 pN at common experimental loading rates. These forces, which constitute about half the mechanical strength of weak covalent bonds, make the system one of the strongest bimolecular interactions described in literature.

Through a combination of single molecule force spectroscopy and steered molecular dynamics simulations it was found that the complex uses a combination of elaborate molecular mechanisms to summon this exceptional mechanical strength. A subdomain adjacent to the actual binding domain termed X module acts as a mechanical effector that actively stabilizes the binding interface. The identification of this indirect stabilization effect is a prime example for the power of single molecule manipulation techniques as it is hard to assess through bulk assays. Moreover, the intermolecular contact area increases as the complex is subjected to external forces. This is achieved through rearrangement of amino acid side chains at the interface, which ultimately protects a common hydrophobic binding patch from attack through solvent molecules. Building on these findings, a novel analysis method for molecular dynamics trajectories was developed to highlight pathways along which external forces are propagated through a molecular complex. This combination of thermodynamic fluctuation theory and network based correlation analysis revealed, that external forces are propagated across the binding interface along angles largely perpendicular to the unbinding direction. This results in a lower net contribution to the amount of work needed to mechanically dissociate the complex.

This framework is expected to provide a platform for exciting future studies, as it remains an open question whether the aforementioned mechanisms for mechanostability, especially the ability to locally redirect external forces in favorable ways, is unique to the investigated system or a more common mechanism. On a more practical note, the extreme mechanical stability combined with easy recombinant expression make the investigated receptor ligand complex a versatile tool for single molecule force spectroscopy applications. It can be used as a molecular anchor to study the unfolding of weaker domains, minimizing potential biasing

effects as discussed in Chapter 5. Further, applications in protein based single molecule cut and paste [228] are conceivable.

Additionally, the theoretical framework depicted in Chapter 5 may help to accurately characterize different mechanical stability in hard to discern dual binding modes as exhibited by type I cohesin dockerin interactions (see Appendix E).

On a more general note, the results presented in this work highlight the great potential of combined single molecule experiments and molecular dynamics simulations. With continuing technical advances on both the experimental and computational side (*e.g.*, to bridge the timescales between experiment and simulation), future studies will undoubtedly contribute to a deeper understanding of the physical concepts governing complex biological functions.

Part III

Appendix

Appendix A

Review Article in the Journal of Structural Biology

Single-Molecule Force Spectroscopy on polyproteins and receptor-ligand complexes: the current toolbox

by

Wolfgang Ott, Markus A. Jobst, Constantin Schoeler, Hermann E. Gaub, and Michael A. Nash

published in

Journal of Structural Biology doi: 10.1016/j.jsb.2016.02.011 (2016)
Reprinted from Ott *et al.* [7] with permission from Elsevier Inc.

The publication in this Appendix was used as a basis for Chapter 1

ARTICLE IN PRESS

Journal of Structural Biology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox

Wolfgang Ott [a,b,1], Markus A. Jobst [a,1], Constantin Schoeler [a,1], Hermann E. Gaub [a], Michael A. Nash [a,c,d,*]

[a] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, 81377 Munich, Germany
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[d] Department of Biosystems Science and Engineering, ETH-Zürich, 4058 Basel, Switzerland

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form 8 February 2016
Accepted 9 February 2016
Available online xxxx

Keywords:
Atomic force microscopy
Single-molecule force spectroscopy
Molecular recognition
Biophysics

ABSTRACT

Single-molecule force spectroscopy sheds light onto the free energy landscapes governing protein folding and molecular recognition. Since only a single molecule or single molecular complex is probed at any given point in time, the technique is capable of identifying low-probability conformations within a large ensemble of possibilities. It furthermore allows choosing certain unbinding pathways through careful selection of the points at which the force acts on the protein or molecular complex. This review focuses on recent innovations in construct design, site-specific bioconjugation, measurement techniques, instrumental advances, and data analysis methods for improving workflow, throughput, and data yield of AFM-based single-molecule force spectroscopy experiments. Current trends that we highlight include customized fingerprint domains, peptide tags for site-specific covalent surface attachment, and polyproteins that are formed through mechanostable receptor–ligand interactions. Recent methods to improve measurement stability, signal-to-noise ratio, and force precision are presented, and theoretical considerations, analysis methods, and algorithms for analyzing large numbers of force–extension curves are further discussed. The various innovations identified here will serve as a starting point to researchers in the field looking for opportunities to push the limits of the technique further.

1. Introduction

The field began in earnest with the introduction of fluid cells for the (at that time) newly developed atomic force microscope (AFM) (Drake et al., 1989). The early 1990s then saw an explosion of the bio-AFM field, which opened the door to high-resolution imaging of proteins and cell surfaces under near-native conditions (Müller et al., 1995; Radmacher et al., 1996, 1992). Shortly thereafter came the realization that individual proteins and DNA molecules, or single receptor–ligand complexes, could be probed with the help of nano- to microscale force transducers (e.g., cantilevers, optically trapped beads, magnetically trapped beads) (Block et al., 1990; Florin et al., 1995; Lee et al., 1994a,b; Smith et al., 1992; Svoboda et al., 1993). It was furthermore discovered that natural polyproteins (e.g., Titin) with repetitive multi-domain structures provided regularly repeating saw-tooth like features in force extension data (Rief et al., 1997a). Artificial (i.e., recombinant) polyproteins quickly came into fashion as internal molecular controls for investigating mechanical properties of protein domains of interest. Since then, engineering of polyproteins has provided a wealth of information about mechanostable motifs in protein folds (Carrion-Vazquez et al., 1999; Oberhauser et al., 1998; Oesterhelt et al., 2000), directional dependence of protein mechanostability (Brockwell et al., 2003; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Kim et al., 2011), and modulation of mechanostability by molecular recognition (Hu and Li, 2014).

Today, force spectroscopy and bio-AFM in general are well established as standard tools in the nanobiosciences, and are regularly used for investigating cell adhesion and cell surface properties (Helenius et al., 2008; Müller et al., 2009; Preiner et al., 2014; Tsukasaki et al., 2007; Wildling et al., 2012), interrogating membrane proteins (Beedle et al., 2015b; Janovjak et al., 2004; Müller, 2008; Müller and Engel, 2007), and measuring mechanical properties of proteins (Beedle et al., 2015a; Bu et al., 2012; Cao et al., 2011; del Rio et al., 2009; Geisler et al., 2010), polysaccharides (Kocun et al., 2011; Rief et al., 1997b) and DNA (Albrecht et al., 2003). Recent studies have already begun to characterize membrane proteins *in vivo* by probing their response to external forces on native living cells (Alsteens et al., 2010; Pfreundschuh et al.,

* Corresponding author.
E-mail address: michael.nash@lmu.de (M.A. Nash).
[1] Theses authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jsb.2016.02.011
1047-8477/

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

2015). There are a number of review articles that thoroughly cover the field from the early years (Carvalho et al., 2013; Casuso et al., 2011; Hoffmann and Dougan, 2012; Lee et al., 2007; Li and Cao, 2010; Marszalek and Dufrêne, 2012; Müller and Dufrêne, 2008; Neuman and Nagy, 2008; Noy, 2011; Rief and Grubmüller, 2002; Sirbuly et al., 2015; Woodside and Block, 2014).

Despite the high level of interest and well-developed method of AFM-SMFS (Single Molecule Force Spectroscopy), there have remained several limitations to the technique that prevent researchers from fully taking advantage of mechano-phenotyping of molecules and cell surfaces. Specifically, low experimental throughput and low yield of useable single-molecule interaction curves have both hampered the widespread adoption of the method, and its application for studying a large number of proteins. The purpose of this review is to highlight recent developments in bioconjugate chemistry, instrumentation, and data processing/algorithms which aim at improving the design process, yield, measurement quality and throughput of AFM-SMFS experiments.

2. Unfolding fingerprints

In typical AFM-SMFS experiments, many thousand force–extension curves are recorded, but only a fraction of these curves contain useable data that describe the behavior of a single molecule. Typically, the majority of curves (~80–99%) contain no interaction, a multiplicity of interactions that are difficult to interpret, or unspecific adhesion events as measurement artifacts. The experimenter is left searching for a needle in a haystack, looking for single-molecule interactions among a vast excess of unusable force–extension curves. In order to filter the data efficiently, the SMFS community has identified a broad range of proteins that can be used as specific identifiers in unfolding traces. We refer to these domains as 'fingerprints' because they provide a unique unfolding step or 'contour-length increment' of defined length that can be used as a filter during data processing. These fingerprint domains are typically globular protein domains with individual unfolding forces and length increments varying across a large range. This ability to choose the length increments and unfolding forces of the fingerprint domains has enabled the design of custom fusion proteins with well-controlled unfolding behaviors. Recent surveys of mechanical properties of different protein domains are provided by Sułkowska and Cieplak (2007), Hoffmann and Dougan (2012).

3. Receptor–ligand SMFS

Protein–protein and protein-small molecule interactions have been widely analyzed with SMFS. Reports of receptor–ligand SMFS include measurements on biotin–avidin (Florin et al., 1994; Lee et al., 1994a,b; Moy et al., 1994; Rico and Moy, 2007; Yuan et al., 2000), antigen–antibody interactions (Hinterdorfer et al., 1996; Morfill et al., 2007; Schwesinger et al., 2000) along with several other protein–protein or small molecule interactions (Lee et al., 2007; Mitchell et al., 2007; Schmidt et al., 2012).

One limitation in the standard method of receptor–ligand SMFS is that the signal lacks single-molecule specificity. Depending on the proteins involved and the experimental conditions (i.e., blocking/passivation steps), and since typically no fingerprint molecules are used, it can be difficult to differentiate non-specific interactions from specific protein-protein recognition. A second limitation of many receptor–ligand SMFS experiments is that pulling geometry is not strictly controlled. While in a standard polyprotein experiment, the force is applied strictly between the N- and C-termini of each domain, coupling of receptors and ligands to AFM tips and substrates is often done through amide linkages formed between amine groups on the proteins and activated NHS-ester groups on the surface or cantilever. This implicates a diversity of pulling geometries which are not strictly controlled, resulting in rupture force distributions that are smeared out or otherwise distorted.

4. Receptor–ligand SMFS with fingerprints

Our group has worked on improving the technique for receptor–ligand SMFS out of sheer necessity (Fig. 1). We were interested in studying a family of receptor–ligand proteins (i.e., cohesin–dockerin, Coh–Doc) involved in carbohydrate recognition and degradation by anaerobic bacteria (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012). These protein receptor–ligand complexes are responsible for building up large extracellular networks of structural scaffold proteins and enzymes. They are linked into these structural networks in well-defined and known orientations (e.g., N-terminal or C-terminal anchoring points). It is important to note that when pulling apart a receptor–ligand complex consisting of two proteins, there are four possible terminal pulling configurations (i.e., N-N′, N-C′, C-N′, C-C′) (Fig. 1B). Many of the Coh–Doc complexes we are interested in possess a clear 'physiological' pulling configuration found in nature, and 'non-physiological' or 'non-native' configurations. To understand their natural mechanical adaptations giving rise to their remarkable assembly strategy, we sought to characterize the mechanical stability of these receptor–ligand complexes in both their native and non-native loading configurations. We found a way to ensure specific interactions by basically combining two previously separate modes of AFM-SMFS (i.e., on polyproteins and receptor–ligand complexes). We fused the Coh and Doc domains separately to different fingerprint domains, and recombinantly produced each construct as a single fusion protein. The fingerprints serve two purposes: (1) they provide site-specific attachment sites through engineered cysteine residues or peptide ligation tags (see section 5) to strictly control loading geometry; (2) they provide predetermined increments in contour length which allows us to filter the datasets for specific single-molecule interactions (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012).

This configuration yields several advantages: We now have the ability to study mechanical stability of receptor–ligand pairs and unfolding of individual domains (i.e., the fingerprints) in a single-experiment with high yield and specificity, eliminating measurement artifacts. We also have a systematic and straightforward way to probe effects of pulling geometry on receptor–ligand unbinding, and to compare native and non-native pulling configurations. The gene design (i.e., N- or C-terminal fingerprint domains) directly reflects the conformation to be investigated. Furthermore, a specific protein domain of interest can now easily be fused to a mechanostable Coh–Doc receptor–ligand pair for characterization. Depending on the expected domain unfolding forces, an appropriately fitting protein receptor–ligand pair can be chosen from a wide range of well-characterized molecules (Table 1). We note that this table does not include every receptor–ligand probed by AFM. For an extensive list of receptor–ligands that were explored with AFM, see Lee et al. (2007). Currently, the mechanically most stable receptor–ligand pair is a Coh–Doc type III complex derived from *R. flavefaciens*, with loading-rate dependent rupture forces between 600 and 800 pN (Schoeler et al., 2015, 2014). Another interaction in a similar force range is the trimeric titin–telethonin complex described by Bertz et al. (2009).

5. Site-specific bioconjugation

Many polyprotein experiments rely on non-specific adsorption of polyproteins onto surfaces (e.g., mica, gold). Receptor–ligand

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

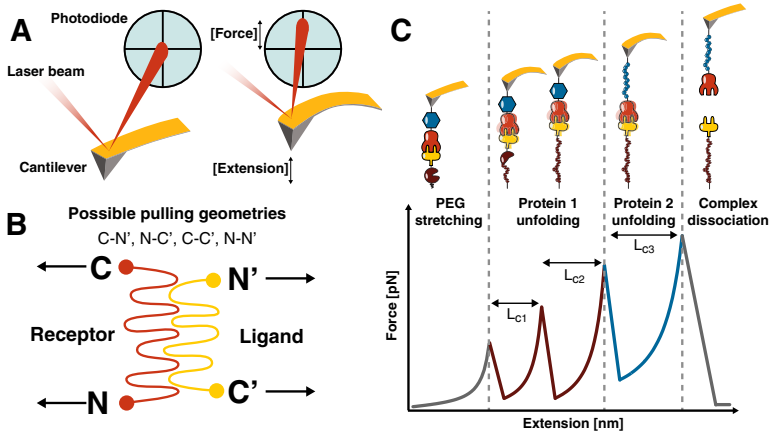


Fig. 1. Configuration for performing receptor–ligand SMFS with (poly)protein fingerprints. (A) Schematic of the measurement setup. The change of force is detected via the differential signal of the laser beam deflection on a quadrant photodiode. (B) For a protein complex consisting of two domains, 4 terminal pulling configurations are possible (N-N′, N-C′, C-N′, C-C′). (C) Fingerprints (brown and blue) are site-specifically and covalently attached to the cantilever and surface. Receptor (orange) and ligand (yellow) form a stable receptor–ligand complex. Note that the fingerprints can be individual sub-domains, or repetitive polyproteins in their own right. Shown is a typical force-extension trace with unfolding of the fingerprints, followed by rupture of the receptor ligand complex. In order to observe unfolding of the fingerprints in sufficient numbers, their most probable unfolding force should lie well below the most probable rupture force of the complex for the given loading rate.

Table 1
Overview of selected receptor–ligand pairs usable as specific handles for protein-based SMFS experiments. Rupture forces depend on immobilization sites for surface conjugation. Note that rupture forces can also vary depending on probe spring constants and loading rates. Abbreviations: NHS: N-hydroxysuccinimide; PEG: poly(ethylene glycol); Mal: maleimide; Cys: cysteine; CoA: coenzyme A; SFP: 4′-phosphopantetheinyl transferase; ybbR-Tag: peptide sequence DSLEFIASKLA; LF: low force unbinding path; HF: high force unbinding path. For the column 'immobilization method', the terminology X (Y) Z means: molecule X is attached to Z mediated by enzyme Y.

Protein handles Handle A:Handle B	Sizes (kDa)	Dissociation force (pN)	Immobilization method	Handle position (N/C)	References
Cohesin:dockerin I	15.4/8.3	122 ± 18.5	NHS-PEG5000-Mal/Cys	C:C	Stahl et al. (2012)
Cohesin:dockerin III	21.6/26.2	606 ± 54	NHS-PEG5000-Mal/Cys	N:C	Schoeler et al. (2015)
		111 ± 30 (LF)	NHS-PEG5000-Mal/CoA (SFP) ybbR	C:C	Schoeler et al. (2015)
		597 ± 67 (HF)	NHS-PEG5000-Mal/CoA (SFP) ybbR		
NiNTA:HIS6	0.2/0.8	153 ± 57	Gold-Cys	n.a.	Verbelen et al. (2007)
Avidin:biotin	66-69/0.2	160 ± 20	Biotinylated BSA	n.a.	Florin et al. (1994)
StrepTagII:streptavidin	1.1/52.8	253 ± 20	BSA/NHS-biotin	n.a.	Wong et al. (1999)
Streptavidin:biotin	52.8/0.2	200	Biotinylated BSA	n.a.	Rico and Moy (2007)
Calmodulin:CBP	16.7/1.1	16.5 ± 1.8	Pulldown via NI-NTA	n.a.	Junker and Rief (2009)
StrepTagII:mono-streptactin	1.1/58.4	116	NHS-PEG5000-Mal/Cys	C:C	Baumann et al. (2015)
		46	NHS-PEG5000-Mal/CoA (SFP) ybbR	N:C	
Anti-GCN4 sFv:GCN4(7P14P)	26.7/4.0	70	NHS-PEG5000-Mal/Cys	N:C	Morfill et al. (2007)
Anti-digoxigenin:digoxigenin	170/0.4	40	NHS-PEG6000	n.a.	Neuert et al. (2006)

AFM-SMFS, however, requires covalent immobilization of the two binding partners to the cantilever and surface, respectively, in order to avoid clogging of the molecules on the cantilever tip. Site-specific (i.e., residue specific) conjugation methods provide strict control over the pulling geometry and result in higher accuracy, precision and reproducibility, compared to conjugation methods resulting in a multiplicity of possible linkage sites (e.g., amine-targeting). Fig. 2 provides an overview of established surface chemistry strategies.

Another advantage of our modular system is the ability to use one construct (i.e., fingerprints with immobilization site) in all desired biochemical or biophysical assays, since immobilization relies on a PEG derivative, which is orthogonal to conventional specific pull down methods. It is compatible with a wide range of binding assays like Western Blotting, ITC, SPR, and ELISA.

The Ni-NTA:HIS_6-tag interaction can be used as force probe as well. This interaction has been employed as an adhesion sensor by probing a cell surface containing His-tagged protein. Since the His-tag is only located at one of the protein's termini, the insertion direction of the protein as well as it's position can be detected (Alsteens et al., 2013; Dupres et al., 2009; Pfreundschuh et al., 2015). This technique is especially useful since the His-tag can be used as a protein purification tag and simultaneously provides a single-molecule force handle.

5.1. Cysteines

Cysteines are relatively rare in proteins, making them attractive as a point mutation residue. The thiol side chain of cysteine is nucleophilic, and will spontaneously react with maleimide leaving groups at neutral pH. It can be used to site-specifically attach proteins to PEG coated surfaces for receptor–ligand AFM-SMFS. Alternatively, engineered cysteines can also be used as oligomerization sites to create disulfide-linked polyproteins, as was done for green fluorescent protein (GFP) (Dietz and Rief, 2006). However, cysteine/thiol-based protein conjugation has some drawbacks, including the tendency of cysteine-modified proteins to multimerize and ultimately aggregate over time, and incompatibility with proteins dis-

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

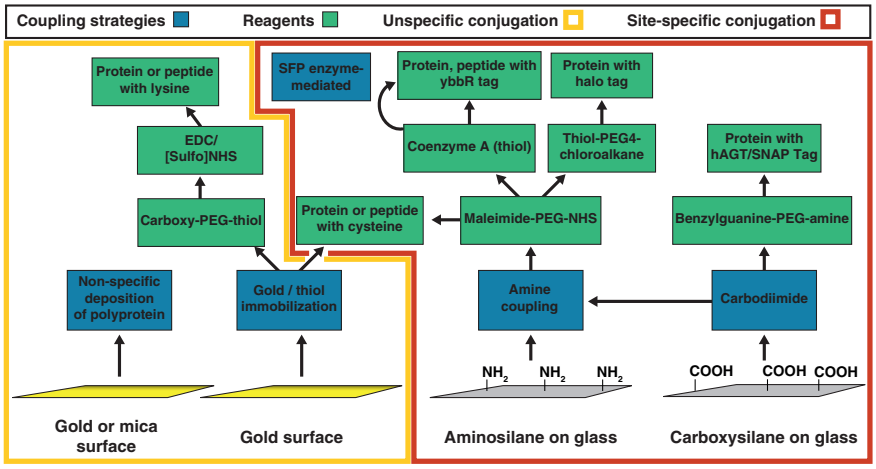


Fig. 2. Surface chemistry and bioconjugation strategies for single-molecule force spectroscopy. The diagram is by no means exhaustive and is roughly divided into site-specific conjugation methods that provide a single anchoring point for proteins to surfaces/cantilevers (right), and unspecific conjugation methods that provide a heterogeneity of loading configurations (i.e., a multiplicity of pulling points) (left).

playing cysteines on their surfaces in their wild-type form. Hence several other conjugation strategies were developed to overcome this challenge. Most of the newer techniques rely on N- or C-terminal attachment sites because the length of the requisite peptide tags or fusion domains makes inclusion into internal sites of a folded protein domain more challenging.

5.2. HaloTag

The active site of the haloalkane dehydrogenase (HaloTag) has been used to covalently immobilize proteins on chloroalkane surfaces. The unfolding forces of the HaloTag depend on its loading geometry (N-terminus: 131 pN; C-terminus: 491 pN). The domain provides an unfolding fingerprint of defined contour length, which also depends on the pulling geometry (N: 66 nm, C: 26.5 nm) (Popa et al., 2013).

5.3. hAGT/SNAP tag

The DNA repair protein O^6-alkylguanine–DNA-alkyltransferase (hAGT, SNAP-tag) binds benzylguanine covalently as a substrate, which can be attached to glass surfaces via an amino-polyethylene glycol (Kufer et al., 2005). With 22 kDa, the SNAP-tag is slightly smaller compared to the HaloTag (34 kDa).

5.4. SpyTag/Catcher

The versatile SpyTag/Catcher system can also be employed for site-specific surface immobilization. The linkage between SpyTag and Catcher is based on an internal protein interaction, which forms an isopeptide (covalent) bond. Based on this observation, the interaction was further developed and engineered, and now consists of a 13 amino acid large SpyTag and the binding domain Spy Catcher (Zakeri et al., 2012).

5.5. ybbR/SFP

The ybbR-Tag is an 11 amino acid protein sequence that is enzymatically linked to coenzyme A (CoA) by 4′-phosphopantetheinyl transferase (SFP) enzyme (Pippig et al., 2014; Yin et al., 2006; Yin et al., 2005). Both ybbR-Tag and the SpyTag/Catcher system have been shown to be N- and C-terminally active. Both tags can also be inserted internally, if the structure of the protein allows it, however, proper folding is not guaranteed and must be evaluated on a case-by-case basis.

5.6. Surface chemistry

Like the modular design of fingerprints and site-specific immobilization tags, surface chemistry can also be modularized to improve workflow. We note that the type of surface chemistry goes hand in hand with the design of the bioconjugation tags for protein production. Our standard approach follows the protocol described by Zimmermann et al. (2010): amino-silanized glass slides and cantilevers are functionalized with a hetero-bifunctional poly (ethylene glycol) (PEG) polymer with an N-hydroxysuccinimide group and a maleimide group at opposing ends. PEG coating provides a passivated surface that resists nonspecific protein adhesion, reducing background and artifacts during measurement. The entropic elasticity behavior of PEG (i.e., persistence length) is similar, although not equal to that of protein backbones, making it a suitable choice for surface conjugation in AFM-SMFS, without interfering too strongly with data interpretation. The maleimide group can then either be modified with CoA containing an inherent thiol group to proceed with ybbR/SFP chemistry, or alternatively directly be reacted with a protein domain displaying a reduced cysteine residue. The PEG incubation can be modified or extended depending on the requirements of the linker and the end group.

6. Advances in measurement techniques

Current advances in measurement resolution, instrument stability and accessible dynamic ranges open up new opportunities for measurements of biomolecules. Here we highlight recent innovations aimed at improving quality and precision of AFM-SMFS measurements.

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

6.1. Improved time resolution

In general, the timescales relevant for protein un-/folding and the corresponding timescale for thermally induced crossing of energy barriers are not fully detectable by common SMFS techniques, which typically resolve slower than 50 μs. Early on, the importance of developing high-speed AFM imaging and force spectroscopy through miniaturization of cantilevers with high resonance frequencies and low viscous drag coefficients was appreciated (Viani et al., 1999a,b). Nonetheless, only recent studies were able to overcome timescale limitations to observe, for example, extraordinarily slow protein misfolding transitions (~0.5 ms) using optical tweezers (Yu et al., 2015). Furthermore, advanced statistical methods extended optical tweezers SMFS time resolution to the ~10 μs range (Žoldák et al., 2013), and optimization of AFM cantilevers for SMFS has pushed the limit toward resolution on the microsecond timescale (Edwards et al., 2015). These developments allow experimentally accessible ranges to approach the lower limits of fast folding transition dynamics (Chung et al., 2012; Schuler and Hofmann, 2013), resolving short-lived intermediate states and yielding important insights into other fast conformational dynamics.

6.2. Bridging the timescale gap to steered molecular dynamics simulations

Recently, experimental measurements were brought into proximity (Dong and Sahin, 2011; He et al., 2012; Schoeler et al., 2015) or even overlap (Rico et al., 2013) with all atom steered molecular dynamics (SMD) simulations. Depending on the size and thus complexity of the simulated system, it has so far been possible to achieve SMD simulation timescales in the nanosecond to mid-microsecond range (Freddolino et al., 2008; Heymann and Grubmüller, 2001; Lee et al., 2009). Rico *et al.* developed a high speed force spectroscopy AFM based on an Ando-type high speed imaging AFM (Ando et al., 2001), with a high resonance frequency (600 kHz) miniature multilayer piezoelectric actuator (calibrated before each experiment and run in open loop mode), and a short cantilever with a high resonance frequency (550 kHz in liquid), and low viscous damping. This system was used to record protein unfolding data at extremely high speeds. To reduce hydrodynamic drag, the sample surface was tilted against the direction of the movement. With these improvements and data acquisition in the megahertz range, they were able to record meaningful and interpretable data at pulling speeds of up to 4000 μm/s, which is about 2–3 orders of magnitude faster than conventional methods and starts overlapping with the range of SMD simulations (Rico et al., 2013). Despite these successes, care must be taken because underdamped or 'ringing' cantilevers like the ones used here are not in agreement with the basic assumptions of the traditional SMFS framework, but can be improved by custom cantilever optimization procedures at the cost of time resolution (Edwards et al., 2015).

6.3. Long-term stability and force precision

Sophisticated measurements of complex biological systems or single molecules often require extraordinarily stable low-drift instruments, capable of continuous long-term data acquisition to gain sufficient and reliable statistics. Active stabilization techniques were developed to enable routine long-term stability and Ångstrom scale precision at room temperature for optical trap setups: differential sample position was measured and regulated with two independently stabilized and MHz modulated lasers, backscattered on sample and probe, and recorded separately on a single photodiode using lock-in amplifiers (Walder et al., 2015). This method is deemed applicable to surface-based and dual-beam optical traps, magnetic tweezers, AFM setups and optical microscopy, including super-resolution techniques.

AFM cantilever long-term stability and force precision can be increased even further by partially removing the reflective gold coating from the cantilever to dramatically reduce cantilever bending caused by the bimetallic effect (Churnside et al., 2012). Stability and precision improvements, which still retain high measurement bandwidths, enable and improve on picoscale force and sub-nanoscale motion measurements of molecular properties and dynamics in various biological systems. These may include groundbreaking investigations like the observation of single RNA polymerase base pair stepping (Abbondanzieri et al., 2005; Zhou et al., 2013), base pair unwinding of helicases (Cheng et al., 2011) and prion misfolding pathways (Yu et al., 2015, 2012). More details on long-term stability measurements and force precision are covered in the recent review of Edwards and Perkins (2016).

6.4. Mapping molecular recognition events: multiparametric imaging modes

The idea of mapping molecular recognition by simultaneously measuring surface topography and force–extension data ('force volume mapping' or 'affinity imaging') was introduced early (Hinterdorfer et al., 1996; Ludwig et al., 1997), and refined to remarkable temporal and spatial resolution. While these molecular recognition imaging techniques turned out to be a valuable tool for detecting and locating specific binding sites on surfaces, their development into dynamic recognition force imaging (Hinterdorfer and Dufrêne, 2006; Raab et al., 1999; Zhang et al., 2014) greatly increased temporal and spatial resolution, while still yielding information about surface elasticity and adhesion, as well as identifying biomolecules at the same time.

Multiparametric imaging modes can simultaneously detect physical properties of the surface and forces exerted on specific biomolecular binding sites. The AFM cantilever oscillates with amplitudes around 100 nm at sub- or low kilohertz frequencies to measure force–distance data, and simultaneously records image topography and other surface properties at sub- or low hertz line-scanning frequencies. The recorded force and topography data is collected orders of magnitude faster compared to force volume mapping methods, yielding imaging speeds comparable to conventional AFM imaging methods (Alsteens et al., 2012; Pfreundschuh et al., 2014). Another benefit of this method is that a large range of loading rates for receptor–ligand dissociation events can be probed in a single experiment, due to the largely varying cantilever tip velocities. Recently, this method was applied to gain nm-scale resolution imaging data of a G protein-coupled receptor (PAR1) in proteoliposomes while characterizing their ligand-binding energy landscape (Alsteens et al., 2015) from loading rates ranging between 1e3 and 1e6 pN/s, already two orders of magnitude higher than conventional force–distance based SMFS. Another recent study demonstrates the ability of this technique to distinguish two different binding events on opposite sides of engineered PAR1 by their unbinding force, and thereby determine their orientation within the lipid bilayer (Pfreundschuh et al., 2015).

6.5. Lateral force sensors

A slightly different approach developed a T-shaped cantilever (Dong et al., 2009; Dong and Sahin, 2011) to drive it at its flexural resonance frequency (~9 kHz) and record force data from cantilever torsion, resulting in a lateral laser deflection signal that was acquired while imaging the sample in conventional tapping mode. Due to the cantilever's high torsional resonance (~115 kHz), unbinding dynamics could be measured at the

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

microsecond timescale and at extraordinarily high loading rates of up to nearly 1e9 pN/s (Dong and Sahin, 2011), about four orders of magnitude faster than conventional SMFS. Force curves and therefore unbinding events and their corresponding force values could be mapped with high spatial and temporal resolution, while providing AFM images that were simultaneously recorded as surface topography. Mechanical elasticity properties of the substrate were also detected in the phase signal.

7. Theory and data analysis

7.1. The data analysis problem

Technical advances greatly increasing the throughput of AFM-SMFS measurements have made automated data analysis protocols an essential requirement. In practice, researchers face the problem of extracting meaningful single molecule signal from large datasets that contain an abundance of unusable data. The use of well-defined fingerprint domains with known unfolding patterns facilitates this procedure greatly. To avoid tedious and time-consuming manual sorting of thousands of data traces, and potential introduction of bias into the data analysis procedure, algorithms which identify the fingerprint unfolding length increments and classify the data correspondingly have been developed and implemented with success (Bosshart et al., 2012; Jobst et al., 2015; Kuhn et al., 2005; Puchner et al., 2008).

7.2. Polymer elasticity models and contour length transformations

Single molecule force measurements generally only gain access to a protein's extension under a given force. The stochastic nature of domain unfolding or complex dissociation under force as well as the non-linear elastic behavior of the polymer backbone chain makes analysis in force-extension space difficult. The same unfolding event is observed over a range of different positions in force–extension curves for multiple measurement cycles as shown in Fig. 3B i.

From a physicist's point of view, mechanical stretching of an unfolded protein domain is described by polymer elasticity models such as the worm-like chain (WLC) (Bustamante et al., 1994), the freely jointed chain (FJC) (Ortiz and Hadziioannou, 1999), or the freely rotating chain (FRC) model (Livadaru et al., 2003). These models contain the free contour length L of the polymer, including surface tethers and unfolded protein backbone, as a parameter. The free contour length is simply the length of the polypeptide along the contour of the biopolymer chain, given a specific folding state (e.g., Fig. 3A). Under a set of physically relevant constraints (L, x, $F > 0$, $x < L$), these elasticity models provide one-to-one mappings from force–extension space into force-contour length space. The models can be solved for the contour length parameter (Jobst et al., 2013; Puchner et al., 2008), yielding an expression for the contour length as a function of force and extension $L(F,x)$. This function can be used to transform force–extension traces from constant speed or force clamp/ramp experiments into contour length space (Fig. 3B ii). The calculated contour length then can be binned (Fig. 3B iii), aligned, and subsequently averaged to precisely locate energy barriers (Fig. 3B iv) along a protein's unfolding pathway, and to classify data sets based on unfolding patterns. This idea was first proposed by Puchner et al. (2008) and has been successfully applied in multiple AFM-SMFS studies (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2014; Stahl et al., 2012; Thoma et al., 2015).

7.3. Worm-like chains, freely-rotating chains and beyond

The WLC model accurately describes a protein's stretching response for forces up to approximately 150 pN. While many protein unfolding or dissociation events take place well within this force regime, some interactions like titin Ig domain unfolding (Rief et al., 1997a), cohesin unfolding (Valbuena et al., 2009), disso-

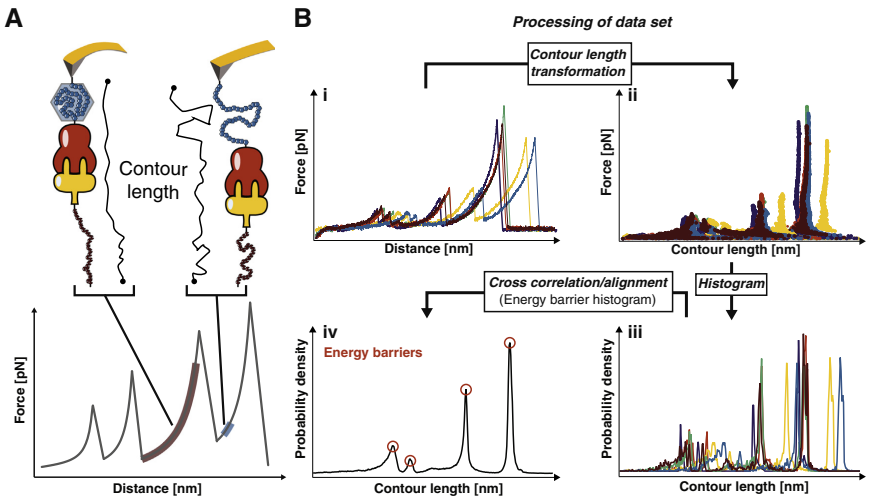


Fig. 3. Assembly of contour length histograms for screening AFM-SMFS datasets. (A) Force–extension traces are transformed into contour length space using an appropriate polymer elasticity model. The choice of the model depends on the force range. (B) Following transformation, the data (i) are plotted in force-contour length space (ii). Force and contour length thresholds are applied and the data are histogrammed (projected onto contour length axis) with an appropriate bin width, i.e., nanometer scale, to obtain the diagram in (iii). Each trace analyzed this way can be searched for a specific contour length increment (distance between two peaks in the probability density vs. contour length plot) corresponding to one of the fingerprints. To obtain a master histogram describing all the observed increments in a dataset, individual histograms reflecting a specific unfolding pathway are aligned by cross-correlation and offsetting along the contour-length axis (iv).

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

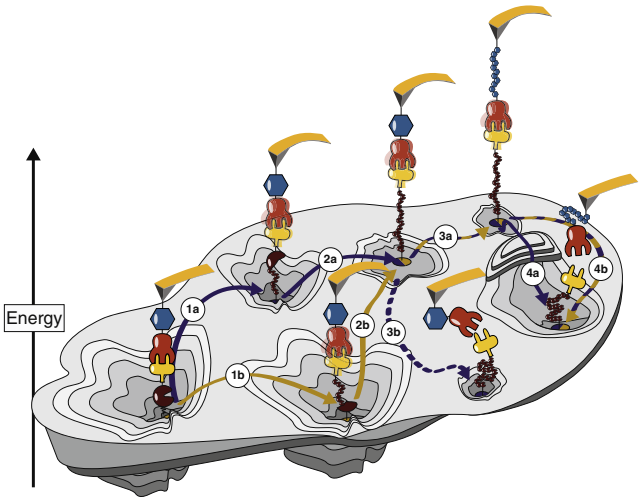


Fig. 4. Schematic depiction of an (un)folding energy landscape. The bound state of a protein receptor–ligand complex can be thought of as a Brownian particle confined to a complex multidimensional energy landscape. At equilibrium, the system can escape the bound state driven by thermal fluctuations. This escape can occur along any pathway on the energy landscape. When measuring the thermal off-rate with bulk assays such as surface plasmon resonance biosensors, a weighted average of all thermally accessible pathways is obtained. In a single-molecule pulling experiment, however, a small subset of pathways is selected, which is defined by the projection of the energy landscape onto the pulling coordinate as illustrated by paths 1–3. Caution is required when comparing data obtained from single molecule techniques with bulk data. In cases where SMFS probes a steep pathway with a high free energy barrier, the fitted zero-force off rate may greatly differ from values obtained by bulk techniques. Path 4 illustrates the thermal escape (4b) versus the forced pathway across an additional energy barrier (4a) by the AFM cantilever.

ciation of skeletal muscle titin–telethonin bonds (Bertz et al., 2009) or dissociation of cellulosomal adhesion complexes (Schoeler et al., 2015, 2014) exhibit much higher unfolding or rupture forces. To adequately describe the elastic response of polymers in such high force regimes, models beyond the standard WLC are required. To address this shortcoming, Hugel et al. (2005) developed quantum mechanical corrections for polymer elasticity models to account for polypeptide backbone stretching at high forces. These corrections can be applied to obtain the contour length at zero force L_0 (Puchner et al., 2008).

Livadaru et al. proposed a more sophisticated model exhibiting three distinct regimes for a protein's stretching response as a function of the applied force (Livadaru et al., 2003). For AFM based SMFS, however, mainly the medium to high force regimes are relevant. The medium force regime of protein stretching, roughly between 10 and 125 pN, exhibits classical WLC stretching behavior, whereas the high force regime shows the behavior of a discrete chain, where the stretching response is independent of the persistence length. This model is most suitable for studying high force interactions, especially when combined with the aforementioned quantum mechanical corrections for backbone stretching.

8. Kinetic and energetic parameters

In dynamic force spectroscopy of receptor–ligand pairs, kinetic and energetic parameters of the complex are of interest. The method most prominently used to extract this information from SMFS experiments is to vary the loading rate by measuring the rupture forces at different pulling speeds in constant speed mode (Baumann et al., 2015; Schoeler et al., 2014; Stahl et al., 2012), or with different slopes in force ramp mode (Oberhauser et al., 2001). The obtained rupture force data are then assembled into a dynamic force spectrum, a plot of most probable rupture forces against their corresponding loading rates. In their comprehensive guide to analysis of SMFS data sets, Noy and Friddle (2013) explain the basic physics of bond stretching. An SMFS measurement corresponds to the stretching of multiple elastic components in series, including the projection of the bond potential onto the pulling axis, the cantilever modeled as a harmonic spring and potential linker molecules with nonlinear elasticity deviating from those under investigation. Such a scenario gives rise to bound and unbound states separated by free energy barriers. By pulling on the harmonic spring, this energy landscape is constantly modulated. Since thermal fluctuations are orders of magnitude faster than changes in the external force, the transition from a bound to an unbound state is thermally driven in common loading rate regimes, as described by Bell (1978), Evans and Ritchie (1997), Izrailev et al. (1997). These models describe a linear dependence of the rupture force on the natural logarithm of the loading rate and give access to the zero force off rate k_0 (exponentially amplified under force) and the distance to the transition state Δx. Theoreticians extended this framework and accounted for modulation of Δx by the applied force (Dudko et al., 2006), and the possibility of rebinding at slow loading rates (Friddle et al., 2012). These newer models predict a nonlinear dependence of the most probable rupture force on the loading rate and give the height of the free energy barrier to unbinding ΔG as an additional parameter. Such non-linear trends were observed experimentally, and a comprehensive list of such data sets is given in Friddle et al. (2012). Joint experimental and computational data sets were also analyzed in recent studies (Rico et al., 2013; Schoeler et al., 2015). As Noy and Friddle (2013) point out, these models should only be used if the force spectrum of interest indeed exhibits a non-linear trend. If this is not given, fitting non-linear models results in non-meaningful fit parameters and the phenomenological model should be used instead.

Although in both bulk measurements and single molecule force measurements at common loading rates, the unbinding process is

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

thermally driven, caution is required when comparing their data. While at unbiased equilibrium, all thermally accessible pathways from the bound state are sampled and the off rate is consequently measured as a weighted average, single molecule force measurements select only a small subset of these pathways due to the defined pulling geometry, as illustrated by paths 1–3 in Fig. 4. In cases where the energy landscape is highly asymmetric and the pulling experiment probes a steep pathway, the off rates obtained from single molecule *vs.* bulk measurements might differ greatly (see Fig. 4, paths 4a vs 4b).

9. Summary and outlook

We highlighted recent advances in experimental design, molecular design, sample preparation, measurement and analysis methods for AFM-SMFS on polyproteins and receptor–ligand complexes. We summarized site-specific bioconjugation strategies to obtain well-defined pulling geometries for improved reliability and reproducibility of experiments. We also highlighted receptor–ligand pairs with high mechanical strength (e.g., cohesin–dockerin), and their application as specific pulling handles in AFM-SMFS for improving experimental throughput and curve yield. Finally, we touched on recent innovations in positional control and cantilever microfabrication for improving time and force resolution and stability of the measurement, on emerging techniques for mapping force responses of surfaces to their topologies, and we discussed theoretical considerations for analyzing large numbers of curves.

In the future, there remain several technical challenges that need to be addressed. One of the limitations of AFM is that it covers a relatively high force range, yet there exist a multitude of biological interactions in the low-force regime that are of interest. Further technical advances in instrument design, cantilever fabrication, and feedback control might further improve force resolution and thereby enable such experiments. A second area for improvement involves sample throughput and parallel screening. With the development of more elaborate, sophisticated and well defined surface immobilization strategies and protein handles, significant gains in throughput can be envisioned. Innovations of the chemistry in combination with efficient data analysis protocols and state of the art instrumentation may pave the way towards in depth study of complex, multi-domain protein systems.

These advances in experimental design and throughput would greatly benefit from refined theoretical frameworks that account for parameters such as cantilever stiffness and ringing whilst maintaining analytical tractability. Consequently, with improved methodology we anticipate the community will be able to address an even wider range of questions about mechanical adaptations of proteins and protein complexes in the future.

Acknowledgments

The authors gratefully acknowledge funding from an advanced Grant to HEG from the European Research Council (Cellufuel Grant 294438), and from a Society in Science – Branco Weiss Fellowship administered by ETH Zürich, Switzerland to MAN. We are also thankful for the financial support of the Deutsche Forschungsgemeinschaft SFB 863.

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., Block, S.M., 2005. Direct observation of base-pair stepping by RNA polymerase. Nature 438, 460–465. http://dx.doi.org/10.1038/nature04268.

Albrecht, C., Blank, K., Lalic-Multhaler, M., Hirler, S., Mai, T., Gilbert, I., Schiffmann, S., Bayer, T., Clausen-Schaumann, H., Gaub, H.E., 2003. DNA: a programmable force sensor. Science 301, 367–370. http://dx.doi.org/10.1126/science.1084713.

Alsteens, D., Garcia, M.C., Lipke, P.N., Dufrêne, Y.F., 2010. Force-induced formation and propagation of adhesion nanodomains in living fungal cells. Proc. Natl. Acad. Sci. USA 107, 20744–20749. http://dx.doi.org/10.1073/pnas.1013893107.

Alsteens, D., Dupres, V., Yunus, S., Latgé, J.-P., Heinisch, J.J., Dufrêne, Y.F., 2012. High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy. Langmuir 28, 16738–16744. http://dx.doi.org/10.1021/la303891j.

Alsteens, D., Trabelsi, H., Soumillion, P., Dufrêne, Y.F., 2013. Multiparametric atomic force microscopy imaging of single bacteriophages extruding from living bacteria. Nat. Commun. 4, 1–7. http://dx.doi.org/10.1038/ncomms3926.

Alsteens, D., Pfreundschuh, M., Zhang, C., Spoerri, P.M., Coughlin, S.R., Kobilka, B.K., Müller, D.J., 2015. Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape. Nat. Methods 12, 845–851. http://dx.doi.org/10.1038/nmeth.3479.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K., Toda, A., 2001. A high-speed atomic force microscope for studying biological macromolecules. Proc. Natl. Acad. Sci. USA 98, 12468–12472. http://dx.doi.org/10.1073/pnas.211400898.

Baumann, F., Bauer, M.S., Milles, L.F., Alexandrovich, A., Gaub, H.E., Pippig, D.A., 2015. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat. Nanotechnol. 1–7. http://dx.doi.org/10.1038/nnano.2015.231.

Beedle, A.E.M., Lezamiz, A., Stirnemann, G., Garcia-Manyes, S., 2015a. The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. Nat. Commun. 6, 1–9. http://dx.doi.org/10.1038/ncomms8894.

Beedle, A.E.M., Williams, A., Relat-Goberna, J., Garcia-Manyes, S., 2015b. Mechanobiology – chemical origin of membrane mechanical resistance and force-dependent signaling. Curr. Opin. Chem. Biol. 29, 87–93. http://dx.doi.org/10.1016/j.cbpa.2015.09.019.

Bell, G.I., 1978. Models for the specific adhesion of cells to cells. Science 200, 618–627. http://dx.doi.org/10.1126/science.347575.

Bertz, M., Wilmanns, M., Rief, M., 2009. The titin–telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. Proc. Natl. Acad. Sci. USA 106, 13307–13310. http://dx.doi.org/10.1073/pnas.0902312106.

Block, S.M., Goldstein, L., Schnapp, B.J., 1990. Bead movement by single kinesin molecules studied with optical tweezers. Nature 348, 348–352. http://dx.doi.org/10.1038/348348a0.

Bosshart, P.D., Frederix, P.L.T.M., Engel, A., 2012. Reference-free alignment and sorting of single-molecule force spectroscopy data. Biophys. J. 102, 2202–2211. http://dx.doi.org/10.1016/j.bpj.2012.03.027.

Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N., Radford, S.E., 2003. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nat. Struct. Biol. 10, 731–737. http://dx.doi.org/10.1038/nsb968.

Bu, T., Wang, H.-C.E., Li, H., 2012. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. Langmuir 28, 12319–12325. http://dx.doi.org/10.1021/la301940g.

Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S., 1994. Entropic elasticity of lambda-phage DNA. Science 265, 1599–1600. http://dx.doi.org/10.1126/science.8079175.

Cao, Y., Li, Y.D., Li, H., 2011. Enhancing the mechanical stability of proteins through a cocktail approach. Biophys. J. 100, 1794–1799. http://dx.doi.org/10.1016/j.bpj.2011.02.030.

Carrion-Vazquez, M., Oberhauser, A.F., Fowler, S.B., Marszalek, P.E., Broedel, S.E., Clarke, J., Fernandez, J.M., 1999. Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl. Acad. Sci. USA 96, 3694–3699. http://dx.doi.org/10.1073/pnas.96.7.3694.

Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P.E., Oberhauser, A.F., Fernandez, J.M., 2003. The mechanical stability of ubiquitin is linkage dependent. Nat. Struct. Biol. 10, 738–743. http://dx.doi.org/10.1038/nsb965.

Carvalho, F.A., Martins, I.C., Santos, N.C., 2013. Atomic force microscopy and force spectroscopy on the assessment of protein folding and functionality. Arch. Biochem. Biophys. 531, 116–127. http://dx.doi.org/10.1016/j.abb.2012.11.007.

Casuso, I., Rico, F., Scheuring, S., 2011. Biological AFM: where we come from – where we are – where we may go. J. Mol. Recogn. 24, 406–413. http://dx.doi.org/10.1002/jmr.1081.

Cheng, W., Arunajadai, S.G., Moffitt, J.R., Tinoco, I.J., Bustamante, C., 2011. Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. Science 333, 1746–1749. http://dx.doi.org/10.1126/science.1206023.

Chung, H.S., McHale, K., Louis, J.M., Eaton, W.A., 2012. Single-molecule fluorescence experiments determine protein folding transition path times. Science 335, 981–984. http://dx.doi.org/10.1126/science.1215768.

Churnside, A.B., Sullan, R.M.A., Nguyen, D.M., Case, S.O., Bull, M.S., King, G.M., Perkins, T.T., 2012. Routine and timely sub-picoNewton force stability and precision for biological applications of atomic force microscopy. Nano Lett. 12, 3557–3561. http://dx.doi.org/10.1021/nl301166w.

del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., Sheetz, M.P., 2009. Stretching single Talin rod molecules activates vinculin binding. Science 323, 638–641. http://dx.doi.org/10.1126/science.1162912.

Dietz, H., Rief, M., 2006. Protein structure by mechanical triangulation. Proc. Natl. Acad. Sci. USA 103, 1244–1247. http://dx.doi.org/10.1073/pnas.0509217103.

Dietz, H., Berkemeier, F., Bertz, M., Rief, M., 2006. Anisotropic deformation response of single protein molecules. Proc. Natl. Acad. Sci. USA 103, 12724–12728. http://dx.doi.org/10.1073/pnas.0602995103.

Dong, M., Sahin, O., 2011. A nanomechanical interface to rapid single-molecule interactions. Nat. Commun. 2, 1–6. http://dx.doi.org/10.1038/ncomms1246.

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

Dong, M., Husale, S., Sahin, O., 2009. Determination of protein structural flexibility by microsecond force spectroscopy. Nat. Nanotechnol. 4, 514–517. http://dx.doi.org/10.1038/nnano.2009.156.

Drake, B., Prater, C.B., Weisenhorn, A.L., Gould, S.A.C., Albrecht, T.R., Quate, C.F., Cannell, D.S., Hansma, H.G., Hansma, P.K., 1989. Imaging crystals, polymers, and processes in water with the atomic force microscope. Science 243, 1386–1389. http://dx.doi.org/10.1126/science.2928794.

Dudko, O.K., Hummer, G., Szabo, A., 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys. Rev. Lett. 96, 108101. http://dx.doi.org/10.1103/PhysRevLett. 96.108101.

Dupres, V., Alsteens, D., Wilk, S., Hansen, B., Heinisch, J.J., Dufrêne, Y.F., 2009. The yeast Wsc1 cell surface sensor behaves like a nanospring in vivo. Nat. Chem. Biol. 5, 857–862. http://dx.doi.org/10.1038/nchembio.220.

Edwards, D.T., Perkins, T.T., 2016. Optimizing force spectroscopy by modifying commercial cantilevers: improved stability, precision, and temporal resolution. J. Struct. Biol. 1–13. http://dx.doi.org/10.1016/j.jsb.2016.01.009.

Edwards, D.T., Faulk, J.K., Sanders, A.W., Bull, M.S., Walder, R., LeBlanc, M.-A., Sousa, M.C., Perkins, T.T., 2015. Optimizing 1-μs-resolution single-molecule force spectroscopy on a commercial atomic force microscope. Nano Lett. 15, 7091–7098. http://dx.doi.org/10.1021/acs.nanolett.5b03166.

Evans, E., Ritchie, K., 1997. Dynamic strength of molecular adhesion bonds. Biophys. J. 72, 1541–1555. http://dx.doi.org/10.1016/S0006-3495(97)78802-7.

Florin, E.L., Moy, V.T., Gaub, H.E., 1994. Adhesion forces between individual ligand–receptor pairs. Science 264, 415–417. http://dx.doi.org/10.1126/science.8153628.

Florin, E.L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V.T., Gaub, H.E., 1995. Sensing specific molecular-interactions with the atomic-force microscope. Biosens. Bioelectron. 10, 895–901. http://dx.doi.org/10.1016/0956-5663(95)99227-C.

Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K., 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. 94, L75–L77. http://dx.doi.org/10.1529/biophysj.108.131565.

Friddle, R.W., Noy, A., De Yoreo, J.J., 2012. Interpreting the widespread nonlinear force spectra of intermolecular bonds. Proc. Natl. Acad. Sci. USA 109, 13573–13578. http://dx.doi.org/10.1073/pnas.1202946109.

Geisler, M., Xiao, S., Puchner, E.M., Gräter, F., Hugel, T., 2010. Controlling the structure of proteins at surfaces. J. Am. Chem. Soc. 132, 17277–17281. http://dx.doi.org/10.1021/ja107212z.

He, C., Genchev, G.Z., Lu, H., Li, H., 2012. Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. J. Am. Chem. Soc. 134, 10428–10435. http://dx.doi.org/10.1021/ja3003205.

Helenius, J., Heisenberg, C.-P., Gaub, H.E., Müller, D.J., 2008. Single-cell force spectroscopy. J. Cell Sci. 121, 1785–1791. http://dx.doi.org/10.1242/jcs.030999.

Heymann, B., Grubmüller, H., 2001. Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces. Biophys. J. 81, 1295–1313. http://dx.doi.org/10.1016/S0006-3495(01)75787-6.

Hinterdorfer, P., Dufrêne, Y.F., 2006. Detection and localization of single molecular recognition events using atomic force microscopy. Nat. Methods 3, 347–355. http://dx.doi.org/10.1038/nmeth871.

Hinterdorfer, P., Baumgartner, W., Gruber, H.J., Schilcher, K., Schindler, H., 1996. Detection and localization of individual antibody–antigen recognition events by atomic force microscopy. Proc. Natl. Acad. Sci. USA 93, 3477–3481. http://dx.doi.org/10.1073/pnas.93.8.3477.

Hoffmann, T., Dougan, L., 2012. Single molecule force spectroscopy using polyproteins. Chem. Soc. Rev. 41, 4781–4796. http://dx.doi.org/10.1039/c2cs35033e.

Hu, X., Li, H., 2014. Force spectroscopy studies on protein–ligand interactions: a single protein mechanics perspective. FEBS Lett. 588, 3613–3620. http://dx.doi.org/10.1016/j.febslet.2014.04.009.

Hugel, T., Rief, M., Seitz, M., Gaub, H.E., Netz, R.R., 2005. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. Phys. Rev. Lett. http://dx.doi.org/10.1103/PhysRevLett.94.048301.

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., Schulten, K., 1997. Molecular dynamics study of unbinding of the avidin–biotin complex. Biophys. J. 72, 1568–1581. http://dx.doi.org/10.1016/S0006-3495(97)78804-0.

Janovjak, H., Struckmeier, J., Hubain, M., Kedrov, A., Kessler, M., Müller, D.J., 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure 12, 871–879. http://dx.doi.org/10.1016/j.str.2004.03.016.

Jobst, M.A., Schoeler, C., Malinowska, K., Nash, M.A., 2013. Investigating receptor–ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. J. Vis. Exp., e50950 http://dx.doi.org/10.3791/50950.

Jobst, M.A., Milles, L.F., Schoeler, C., Ott, W., Fried, D.B., Bayer, E.A., Gaub, H.E., Nash, M.A., 2015. Resolving dual binding conformations of cellulosome cohesin–dockerin complexes using single-molecule force spectroscopy. Elife 4, 1031. http://dx.doi.org/10.7554/eLife.9.

Junker, J.P., Rief, M., 2009. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. Proc. Natl. Acad. Sci. USA 106, 14361–14366. http://dx.doi.org/10.1073/pnas.0904654106.

Kim, M., Wang, C.-C., Benedetti, F., Rabbi, M., Bennett, V., Marszalek, P.E., 2011. Nanomechanics of streptavidin hubs for molecular materials. Adv. Mater. 23, 5684–5688. http://dx.doi.org/10.1002/adma.201103316.

Kocun, M., Grandbois, M., Cuccia, L.A., 2011. Single molecule atomic force microscopy and force spectroscopy of chitosan. Colloids Surf. B Biointerfaces 82, 470–476. http://dx.doi.org/10.1016/j.colsurfb.2010.10.004.

Kufer, S.K., Dietz, H., Albrecht, C., Blank, K., Kardinal, A., Rief, M., Gaub, H.E., 2005. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. Eur. Biophys. J. 35, 72–78. http://dx.doi.org/10.1007/s00249-005-0010-1.

Kuhn, M., Janovjak, H., Hubain, M., Müller, D.J., 2005. Automated alignment and pattern recognition of single-molecule force spectroscopy data. J. Microscopy 218, 125–132. http://dx.doi.org/10.1111/j.1365-2818.2005.01478.x.

Lee, C.-K., Chrisey, L.A., Colton, R.J., 1994a. Direct measurement of the forces between complementary strands of DNA. Science. http://dx.doi.org/10.1126/science.7973628.

Lee, G.U., Kidwell, D.A., Colton, R.J., 1994b. Sensing discrete streptavidin–biotin interactions with atomic force microscopy. Langmuir. http://dx.doi.org/10.1021/la00014a003.

Lee, C.-K., Wang, Y.-M., Huang, L.-S., Lin, S., 2007. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. Micron 38, 446–461. http://dx.doi.org/10.1016/j.micron.2006.06.014.

Lee, E.H., Hsin, J., Sotomayor, M., Comellas, G., Schulten, K., 2009. Discovery through the computational microscope. Structure 17, 1295–1306. http://dx.doi.org/10.1016/j.str.2009.09.001.

Li, H., Cao, Y., 2010. Protein mechanics: from single molecules to functional biomaterials. Acc. Chem. Res. 43, 1331–1341. http://dx.doi.org/10.1021/ar100057a.

Livadaru, L., Netz, R.R., Kreuzer, H.J., 2003. Stretching response of discrete semiflexible polymers. Macromolecules 36, 3732–3744. http://dx.doi.org/10.1021/ma020751g.

Ludwig, M., Dettmann, W., Gaub, H.E., 1997. Atomic force microscope imaging contrast based on molecular recognition. Biophys. J. 72, 445–448. http://dx.doi.org/10.1016/S0006-3495(97)78685-5.

Marszalek, P.E., Dufrêne, Y.F., 2012. Stretching single polysaccharides and proteins using atomic force microscopy. Chem. Soc. Rev. 41, 3523–3534. http://dx.doi.org/10.1039/c2cs15329g.

Mitchell, G., Lamontagne, C.-A., Lebel, R., Grandbois, M., Malouin, F., 2007. Single-molecule dynamic force spectroscopy of the fibronectin–heparin interaction. Biochem. Biophys. Res. Commun. 364, 595–600. http://dx.doi.org/10.1016/j.bbrc.2007.10.034.

Morfill, J., Blank, K., Zahnd, C., Luginbühl, B., Kühner, F., Gottschalk, K.-E., Plückthun, A., Gaub, H.E., 2007. Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. Biophys. J. 93, 3583–3590. http://dx.doi.org/10.1529/biophysj.107.112532.

Moy, V.T., Florin, E.L., Gaub, H.E., 1994. Intermolecular forces and energies between ligands and receptors. Science 266, 257–259. http://dx.doi.org/10.1126/science.7939660.

Müller, D.J., 2008. AFM: a nanotool in membrane biology. Biochemistry 47, 7986–7998. http://dx.doi.org/10.1021/bi800753x.

Müller, D.J., Dufrêne, Y.F., 2008. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. Nat. Nanotechnol. 3, 261–269. http://dx.doi.org/10.1038/nnano.2008.100.

Müller, D.J., Engel, A., 2007. Atomic force microscopy and spectroscopy of native membrane proteins. Nat. Protocols 2, 2191–2197. http://dx.doi.org/10.1038/nprot.2007.309.

Müller, D.J., Büldt, G., Engel, A., 1995. Force-induced conformational change of bacteriorhodopsin. J. Mol. Biol. 249, 239–243. http://dx.doi.org/10.1006/jmbi.1995.0292.

Müller, D.J., Helenius, J., Alsteens, D., Dufrêne, Y.F., 2009. Force probing surfaces of living cells to molecular resolution. Nat. Chem. Biol. 5, 383–390. http://dx.doi.org/10.1038/nchembio.181.

Neuert, G., Albrecht, C., Pamir, E., Gaub, H.E., 2006. Dynamic force spectroscopy of the digoxigenin–antibody complex. FEBS Lett. 580, 505–509. http://dx.doi.org/10.1016/j.febslet.2005.12.052.

Neuman, K.C., Nagy, A., 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat. Methods 5, 491–505. http://dx.doi.org/10.1038/nmeth.1218.

Noy, A., 2011. Force spectroscopy 101: how to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment. Curr. Opin. Chem. Biol. 15, 710–718. http://dx.doi.org/10.1016/j.cbpa.2011.07.020.

Noy, A., Friddle, R.W., 2013. Practical single molecule force spectroscopy: how to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope. Methods 60, 142–150. http://dx.doi.org/10.1016/j.ymeth.2013.03.014.

Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., Fernandez, J.M., 1998. The molecular elasticity of the extracellular matrix protein tenascin. Nature 393, 181–185. http://dx.doi.org/10.1038/30270.

Oberhauser, A.F., Hansma, P.K., Carrion-Vazquez, M., Fernandez, J.M., 2001. Stepwise unfolding of titin under force-clamp atomic force microscopy. Proc. Natl. Acad. Sci. USA 98, 468–472. http://dx.doi.org/10.1073/pnas.021321798.

Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E., Müller, D.J., 2000. Unfolding pathways of individual bacteriorhodopsins. Science 288, 143–146. http://dx.doi.org/10.1126/science.288.5463.143.

Ortiz, C., Hadziioannou, G., 1999. Entropic elasticity of single polymer chains of poly (methacrylic acid) measured by atomic force microscopy. Macromolecules 32, 780–787. http://dx.doi.org/10.1021/ma981245n.

Otten, M., Ott, W., Jobst, M.A., Milles, L.F., Verdorfer, T., Pippig, D.A., Nash, M.A., Gaub, H.E., 2014. From genes to protein mechanics on a chip. Nat. Methods 11, 1127–1130. http://dx.doi.org/10.1038/nmeth.3099.

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

Pfreundschuh, M., Alsteens, D., Hilbert, M., Steinmetz, M.O., Müller, D.J., 2014. Localizing chemical groups while imaging single native proteins by high-resolution atomic force microscopy. Nano Lett. 14, 2957–2964. http://dx.doi.org/10.1021/nl5012905.

Pfreundschuh, M., Alsteens, D., Wieneke, R., Zhang, C., Coughlin, S.R., Tampé, R., Kobilka, B.K., Müller, D.J., 2015. Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM. Nat. Commun. 6, 1–7. http://dx.doi.org/10.1038/ncomms9857.

Pippig, D.A., Baumann, F., Strackharn, M., Aschenbrenner, D., Gaub, H.E., 2014. Protein–DNA chimeras for nano assembly. ACS Nano 8, 6551–6555. http://dx.doi.org/10.1021/nn501644w.

Popa, I., Berkovich, R., Alegre-Cebollada, J., Badilla, C.L., Rivas-Pardo, J.A., Taniguchi, Y., Kawakami, M., Fernandez, J.M., 2013. Nanomechanics of HaloTag tethers. J. Am. Chem. Soc. 135, 12762–12771. http://dx.doi.org/10.1021/ja4056382.

Preiner, J., Kodera, N., Tang, J., Ebner, A., Brameshuber, M., Blaas, D., Gelbmann, N., Gruber, H.J., Ando, T., Hinterdorfer, P., 2014. IgGs are made for walking on bacterial and viral surfaces. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms5394.

Puchner, E.M., Franzen, G., Gautel, M., Gaub, H.E., 2008. Comparing proteins by their unfolding pattern. Biophys. J. 95, 426–434. http://dx.doi.org/10.1529/biophysj.108.129999.

Raab, A., Hansma, H.G., Badt, D., Smith-Gill, S.J., 1999. Antibody recognition imaging by force microscopy. Nature. http://dx.doi.org/10.1038/12898.

Radmacher, M., Tillmann, R.W., Fritz, M., Gaub, H.E., 1992. From molecules to cells – imaging soft samples with the atomic force microscope. Science 257, 1900–1905. http://dx.doi.org/10.1126/science.1411505.

Radmacher, M., Fritz, M., Kacher, C.M., Cleveland, J.P., Hansma, P.K., 1996. Measuring the viscoelastic properties of human platelets with the atomic force microscope. Biophys. J. 70, 556–567. http://dx.doi.org/10.1016/S0006-3495(96)79602-9.

Rico, F., Moy, V.T., 2007. Energy landscape roughness of the streptavidin–biotin interaction. J. Mol. Recogn. 20, 495–501. http://dx.doi.org/10.1002/jmr.841.

Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M., Scheuring, S., 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. Science 342, 741–743. http://dx.doi.org/10.1126/science.1239764.

Rief, M., Grubmüller, H., 2002. Force spectroscopy of single biomolecules. Chemphyschem 3, 255–261. http://dx.doi.org/10.1002/1439-7641(20020315)3:3<255::AID-CPHC255>3.0.CO;2-M.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., Gaub, H.E., 1997a. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276, 1109–1112. http://dx.doi.org/10.1126/science.276.5315.1109.

Rief, M., Oesterhelt, F., Heymann, B., Gaub, H.E., 1997b. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. Science 275, 1295–1297. http://dx.doi.org/10.1126/science.275.5304.1295.

Schmidt, S.W., Filippov, P., Kersch, A., Beyer, M.K., Clausen-Schaumann, H., 2012. Single-molecule force-clamp experiments reveal kinetics of mechanically activated silyl ester hydrolysis. ACS Nano 6, 1314–1321. http://dx.doi.org/10.1021/nn204111w.

Schoeler, C., Malinowska, K.H., Bernardi, R.C., Milles, L.F., Jobst, M.A., Durner, E., Ott, W., Fried, D.B., Bayer, E.A., Schulten, K., Gaub, H.E., Nash, M.A., 2014. Ultrastable cellulosome–adhesion complex tightens under load. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms6635.

Schoeler, C., Bernardi, R.C., Malinowska, K.H., Durner, E., Ott, W., Bayer, E.A., Schulten, K., Nash, M.A., Gaub, H.E., 2015. Mapping mechanical force propagation through biomolecular complexes. Nano Lett. 15, 7370–7376. http://dx.doi.org/10.1021/acs.nanolett.5b02727.

Schuler, B., Hofmann, H., 2013. Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. Curr. Opin. Struct. Biol. 23, 36–47. http://dx.doi.org/10.1016/j.sbi.2012.10.008.

Schwesinger, F., Ros, R., Strunz, T., Anselmetti, D., Guntherodt, H.-J., Honegger, A., Jermutus, L., Tiefenauer, L., Plückthun, A., 2000. Unbinding forces of single antibody–antigen complexes correlate with their thermal dissociation rates. Proc. Natl. Acad. Sci. USA 97, 9972–9977. http://dx.doi.org/10.1073/pnas.97.18.9972.

Sirbuly, D.J., Friddle, R.W., Villanueva, J., Huang, Q., 2015. Nanomechanical force transducers for biomolecular and intracellular measurements: is there room to shrink and why do it? Rep. Prog. Phys. 1–22. http://dx.doi.org/10.1088/0034-4885/78/2/024101.

Smith, S.B., Finzi, L., Bustamante, C., 1992. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. Science 258, 1122–1126. http://dx.doi.org/10.1126/science.1439819.

Stahl, S.W., Nash, M.A., Fried, D.B., Slutzki, M., Barak, Y., Bayer, E.A., Gaub, H.E., 2012. Single-molecule dissection of the high-affinity cohesin–dockerin complex. Proc. Natl. Acad. Sci. USA 109, 20431–20436. http://dx.doi.org/10.1073/pnas.1211929109.

Sułkowska, J.I., Cieplak, M., 2007. Mechanical stretching of proteins—a theoretical survey of the protein data bank. J. Phys. Condens. Matter 19, 283201. http://dx.doi.org/10.1088/0953-8984/19/28/283201.

Svoboda, K., Schmidt, C.F., Schnapp, B.J., Block, S.M., 1993. Direct observation of kinesin stepping by optical trapping interferometry. Nature 365, 721–727. http://dx.doi.org/10.1038/365721a0.

Thoma, J., Burmann, B.M., Hiller, S., Müller, D.J., 2015. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. Nat. Struct. Mol. Biol. 22, 795–802. http://dx.doi.org/10.1038/nsmb.3087.

Tsukasaki, Y., Kitamura, K., Shimizu, K., Iwane, A.H., Takai, Y., Yanagida, T., 2007. Role of multiple bonds between the single cell adhesion molecules, nectin and cadherin, revealed by high sensitive force measurements. J. Mol. Biol. 367, 996–1006. http://dx.doi.org/10.1016/j.jmb.2006.12.022.

Valbuena, A., Oroz, J., Hervas, R., Manuel Vera, A., Rodriguez, D., Menendez, M., Sulkowska, J.I., Cieplak, M., Carrion-Vazquez, M., 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc. Natl. Acad. Sci. USA 106, 13791–13796. http://dx.doi.org/10.1073/pnas.0813093106.

Verbelen, C., Gruber, H.J., Dufrêne, Y.F., 2007. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. J. Mol. Recogn. 20, 490–494. http://dx.doi.org/10.1002/jmr.833.

Viani, M.B., Schaffer, T.E., Chand, A., Rief, M., Gaub, H.E., Hansma, P.K., 1999a. Small cantilevers for force spectroscopy of single molecules. J. Appl. Phys. 86, 2258–2262. http://dx.doi.org/10.1063/1.371039.

Viani, M.B., Schaffer, T.E., Paloczi, G.T., Pietrasanta, L.I., Smith, B.L., Thompson, J.B., Richter, M., Rief, M., Gaub, H.E., Plaxco, K.W., Cleland, A.N., Hansma, H.G., Hansma, P.K., 1999b. Fast imaging and fast force spectroscopy of single biopolymers with a new atomic force microscope designed for small cantilevers. Rev. Sci. Instrum. 70, 4300–4303. http://dx.doi.org/10.1063/1.1150069.

Walder, R., Paik, D.H., Bull, M.S., Sauer, C., Perkins, T.T., 2015. Ultrastable measurement platform: sub-nm drift over hours in 3D at room temperature. Opt. Express 23, 16554–16564. http://dx.doi.org/10.1364/OE.23.016554.

Wildling, L., Rankl, C., Haselgrübler, T., Gruber, H.J., Holy, M., Newman, A.H., Zou, M.-F., Zhu, R., Freissmuth, M., Sitte, H.H., Hinterdorfer, P., 2012. Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells. J. Biol. Chem. 287, 105–113. http://dx.doi.org/10.1074/jbc.M111.304873.

Wong, J., Chilkoti, A., Moy, V.T., 1999. Direct force measurements of the streptavidin–biotin interaction. Biomol. Eng. 16, 45–55. http://dx.doi.org/10.1016/S1050-3862(99)00035-2.

Woodside, M.T., Block, S.M., 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu. Rev. Biophys. 43, 19–39. http://dx.doi.org/10.1146/annurev-biophys-051013-022754.

Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R., Walsh, C.T., 2005. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc. Natl. Acad. Sci. USA 102, 15815–15820. http://dx.doi.org/10.1073/pnas.0507705102.

Yin, J., Lin, A.J., Golan, D.E., Walsh, C.T., 2006. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. Nat. Protocols 1, 280–285. http://dx.doi.org/10.1038/nprot.2006.43.

Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Sosova, I., Woodside, M.T., 2012. Direct observation of multiple misfolding pathways in a single prion protein molecule. Proc. Natl. Acad. Sci. USA 109, 5283–5288. http://dx.doi.org/10.1073/pnas.1107736109.

Yu, H., Dee, D.R., Liu, X., Brigley, A.M., Sosova, I., Woodside, M.T., 2015. Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. Proc. Natl. Acad. Sci. USA 112, 8308–8313. http://dx.doi.org/10.1073/pnas.1419197112.

Yuan, C., Chen, A., Kolb, P., Moy, V.T., 2000. Energy landscape of streptavidin−biotin complexes measured by atomic force microscopy. Biochemistry 39, 10219–10223. http://dx.doi.org/10.1021/bi992715o.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., Howarth, M., 2012. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc. Natl. Acad. Sci. USA 109, 4347–4348. http://dx.doi.org/10.1073/pnas.1115485109.

Zhang, S., Aslan, H., Besenbacher, F., Dong, M., 2014. Quantitative biomolecular imaging by dynamic nanomechanical mapping. Chem. Soc. Rev. 43, 7412–7429. http://dx.doi.org/10.1039/C4CS00176A.

Zhou, J., Schweikhard, V., Block, S.M., 2013. Single-molecule studies of RNAPII elongation. Biochem. Biophys. Acta 1829, 29–38. http://dx.doi.org/10.1016/j.bbagrm.2012.08.006.

Zimmermann, J.L., Nicolaus, T., Neuert, G., Blank, K., 2010. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. Nat. Protocols 5, 975–985. http://dx.doi.org/10.1038/nprot.2010.49.

Žoldák, G., Stigler, J., Pelz, B., Li, H., 2013. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. Proc. Natl. Acad. Sci. USA 110, 18156–18161. http://dx.doi.org/10.1073/pnas.1311495110.

Please cite this article in press as: Ott, W., et al. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.02.011

Appendix B

Article in Nature Communications Chapter 3 in original form

Ultrastable Cellulosome Adhesion Complex Tightens Under Load

by

Constantin Schoeler, Klara H. Malinowska, Rafael C. Bernardi, Lukas F. Milles, Markus A. Jobst, Ellis Durner, Wolfgang Ott, Daniel B. Fried, Edward A. Bayer, Klaus Schulten, Hermann E. Gaub, and Michael A. Nash

published in

Nature Communications 5:5635 doi:10.1038/ncomms6635 (2014)

nature COMMUNICATIONS

ARTICLE

Received 25 Jun 2014 | Accepted 22 Oct 2014 | Published 8 Dec 2014

DOI: 10.1038/ncomms6635 OPEN

Ultrastable cellulosome-adhesion complex tightens under load

Constantin Schoeler[1,*], Klara H. Malinowska[1,*], Rafael C. Bernardi[2], Lukas F. Milles[1], Markus A. Jobst[1], Ellis Durner[1], Wolfgang Ott[1], Daniel B. Fried[3], Edward A. Bayer[3], Klaus Schulten[2,4], Hermann E. Gaub[1] & Michael A. Nash[1]

Challenging environments have guided nature in the development of ultrastable protein complexes. Specialized bacteria produce discrete multi-component protein networks called cellulosomes to effectively digest lignocellulosic biomass. While network assembly is enabled by protein interactions with commonplace affinities, we show that certain cellulosomal ligand–receptor interactions exhibit extreme resistance to applied force. Here, we characterize the ligand–receptor complex responsible for substrate anchoring in the *Ruminococcus flavefaciens* cellulosome using single-molecule force spectroscopy and steered molecular dynamics simulations. The complex withstands forces of 600–750 pN, making it one of the strongest bimolecular interactions reported, equivalent to half the mechanical strength of a covalent bond. Our findings demonstrate force activation and inter-domain stabilization of the complex, and suggest that certain network components serve as mechanical effectors for maintaining network integrity. This detailed understanding of cellulosomal network components may help in the development of biocatalysts for production of fuels and chemicals from renewable plant-derived biomass.

[1] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany. [2] Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3] Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel. [4] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.A.N. (email: michael.nash@lmu.de).

Cellulosomes are protein networks designed by nature to degrade lignocellulosic biomass[1]. These networks comprise intricate assemblies of conserved subunits including catalytic domains, scaffold proteins, carbohydrate binding modules (CBMs), cohesins (Cohs), dockerins (Docs) and X-modules (XMods) of unknown function. Coh:Doc pairs form complexes with high affinity and specificity[2], and provide connectivity to a myriad of cellulosomal networks with varying Coh:Doc network topology[3–5]. The most intricate cellulosome known to date is produced by *Ruminococcus flavefaciens* (*R.f.*)[6,7] and contains several primary and secondary scaffolds along with over 220 Doc-bearing protein subunits[8].

The importance of cellulolytic enzymes for the production of renewable fuels and chemicals from biomass has highlighted an urgent need for improved fundamental understanding of how cellulosomal networks achieve their impressive catalytic activity[9]. Two of the mechanisms known to increase the catalytic activity of cellulosomes are proximity and targeting effects[10]. Proximity refers to the high local concentration of enzymes afforded by incorporation into nanoscale networks, while targeting refers to specific binding of cellulosomes to substrates. Protein scaffolds and CBM domains are both critical in this context as they mediate interactions between comparatively large bacterial cells and cellulose particles. As many cellulosomal habitats (for example, cow rumen) exhibit strong flow gradients, shear forces will accordingly stress bridging scaffold components mechanically *in vivo*. Protein modules located at stressed positions within these networks should therefore be preselected for high mechanostability. However, thus far very few studies on the mechanics of carbohydrate-active proteins or cellulosomal network components have been reported[11].

In the present study we sought to identify cellulosomal network junctions with maximal mechanical stability. We chose an XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose in the rumen. The complex links the *R. flavefaciens* cell wall to the cellulose substrate via two CBM domains located at the N-terminus of the CttA scaffold, as shown in Fig. 1a. The crystal structure of the complex solved by X-ray crystallography[12] is shown in Fig. 1b. XMod-Doc tandem dyads such as this one are a common feature in cellulosomal networks. Bulk biochemical assays on XMod-Docs have demonstrated that XMods improve Doc solubility and increase biochemical affinity of Doc:Coh complex formation[13]. Crystallographic studies conducted on XMod-Doc:Coh complexes have revealed direct contacts between XMods and their adjacent Docs[12,14]. In addition, many XMods (for example, PDB 2B59, 1EHX, 3PDD) have high β-strand content and fold with N- and C-termini at opposite ends of the molecule, suggestive of robust mechanical clamp motifs at work[15,16]. These observations all suggest a mechanical role for XMods. Here we perform AFM single-molecule force spectroscopy experiments and steered molecular dynamics simulations to understand the mechanostability of the XMod-Doc:Coh cellulosomal ligand–receptor complex. We conclude that the high mechanostability we observe originates from molecular mechanisms, including stabilization of Doc by the adjacent XMod domain and catch bond behaviour that causes the complex to increase in contact area on application of force.

Results and Discussion

Single-molecule experiments. We performed single-molecule force spectroscopy (SMFS) experiments with an atomic force miscroscope (AFM) to probe the mechanical dissociation of XMod-Doc:Coh. Xylanase (Xyn) and CBM fusion domains on the XMod-Doc and Coh modules, respectively, provided identifiable unfolding patterns permitting screening of large data sets of force-distance curves[17–19]. Engineered cysteines and/or peptide tags on the CBM and Xyn marker domains were used to covalently immobilize the binding partners in a site-specific manner to an AFM cantilever or cover glass via poly(ethylene glycol) (PEG) linkers. The pulling configuration with Coh-CBM immobilized on the cantilever is referred to as configuration I, as shown in Fig. 1c. The reverse configuration with Coh-CBM on the cover glass is referred to as configuration II. In a typical

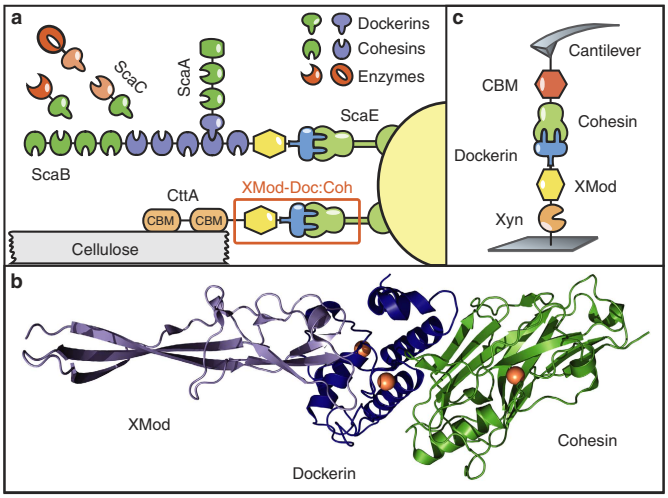


Figure 1 | System overview. (**a**) Schematic of selected components of the *R. flavefaciens* cellulosome. The investigated XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose is highlighted in orange. (**b**) Crystal structure of the XMod-Doc:Coh complex. Ca^{2+} ions are shown as orange spheres. (**c**) Depiction of experimental pulling configuration I, with Coh-CBM attached to the cantilever tip and Xyn-XMod-Doc attached to the glass surface.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms6635 ARTICLE

experimental run we collected about 50,000 force extension traces from a single cantilever. We note that the molecules immobilized on the cantilever and glass surfaces were stable over thousands of pulling cycles.

We sorted the data by first searching for contour length increments that matched our specific xylanase and CBM fingerprint domains. After identifying these specific traces (Fig. 2a), we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history. To assign protein subdomains to the observed unfolding patterns, we transformed the data into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching (QM-FRC, Supplementary Note 1, Supplementary Fig. 1)[20,21]. The fit parameter-free QM-FRC model describes protein stretching at forces >200 pN more accurately than the commonly used worm-like chain (WLC) model[20,22]. The resulting contour length histogram is shown in Fig. 2b. Peak-to-peak distances in the histogram represent contour length increments of unfolded protein domains. Assuming a length per stretched amino acid of 0.365 nm and accounting for the folded length of each subdomain, we compared the observed increments to the polypeptide lengths of individual subdomains of the Xyn-XMod-Doc and Coh-CBM fusion proteins. Details on contour length estimates and domain assignments are shown in Supplementary Table 1.

Unfolding patterns in configuration I showed PEG stretching followed by a three-peaked Xyn fingerprint (Fig. 1a, top trace, green), which added 90 nm of contour length to the system. Xyn unfolding was followed by CBM unfolding at $\sim$150 pN with 55 nm of contour length added. Finally, the XMod-Doc:Coh complex dissociated at an ultra-high rupture force of $\sim$600 pN. The loading rate dependence of the final rupture event for curves of subtype 1 is plotted in Fig. 2c (blue). The measured complex rupture force distributions are shown in Supplementary Fig. 2.

Less frequently (35–40% of traces) we observed a two-step dissociation process wherein the XMod unfolded before Doc:Coh rupture as shown in Fig. 2a (middle trace, orange). In these cases, the final dissociation exhibited a much lower rupture force ($\sim$300 pN) than the preceding XMod unfolding peak, indicating the strengthening effect of XMod was lost, and XMod was no longer able to protect the complex from dissociation at high force. The loading rate dependency of Doc:Coh rupture occurring immediately following XMod unfolding is shown in Fig. 2c (grey).

In configuration II (Fig. 2a, bottom trace), with the Xyn-XMod-Doc attached to the cantilever, the xylanase fingerprint was lost after the first few force extension traces acquired in the data set. This indicated the Xyn domain did not refold within the timescale of the experiment once unfolded, consistent with prior work[17,18]. CBM and XMod unfolding events were observed repeatedly throughout the series of acquired force traces in both configurations I and II, indicating these domains were able to refold while attached to the cantilever over the course of the experiment.

We employed the Bell-Evans model[23] (Supplementary Note 2) to analyse the final rupture of the complex through the effective distance to the transition state (Δx) and the natural off-rate (k_{off}). The fits to the model yielded values of $\Delta x = 0.13$ nm and $k_{off} = 7.3 \times 10^{-7} s^{-1}$ for an intact XMod, and $\Delta x = 0.19$ nm and $k_{off} = 4.7 \times 10^{-4} s^{-1}$ for the 'shielded' rupture following XMod unfolding (Fig. 2c). These values indicate that the distance to the transition state is increased following XMod unfolding, reflecting an overall softening of the binding interface. Distances to the transition state observed for other ligand–receptor pairs are typically on the order of $\sim$0.7 nm (ref. 17). The extremely short Δx of 0.13 nm observed here suggests that mechanical unbinding for this complex is highly coordinated. We further analysed the unfolding of XMod in the Bell-Evans picture and found values of $\Delta x = 0.15$ and $k_{off} = 2.6 \times 10^{-6} s^{-1}$. The loading

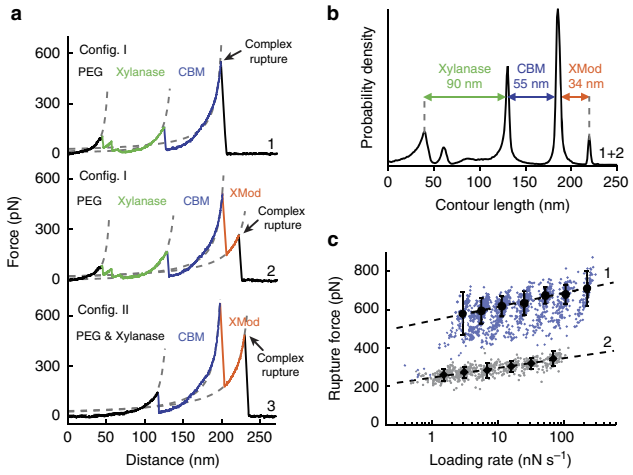


Figure 2 | Experimental SMFS unfolding traces. (**a**) Unfolding fingerprints from pulling configuration I (curves 1 & 2) and configuration II (curve 3). The QM-FRC model (dashed lines) was used to estimate the contour lengths of the unfolded modules. (**b**) Contour length histogram obtained from 127 force extension traces (Config. I). The peak-to-peak increments correspond to Xyn, CBM and XMod amino-acid sequence lengths. (**c**) Dynamic force spectra for the final Doc:Coh complex rupture peaks obtained from 2,122 force-extension traces. The blue points show Doc:Coh ruptures that occurred with an intact XMod, while grey points show ruptures immediately following XMod unfolding. Black circles and diamonds represent the most probable rupture force/loading rate obtained by Gaussian fitting at each pulling speed. Error bars are ± 1 s.d. Dashed lines are least square fits to the Bell-Evans model.

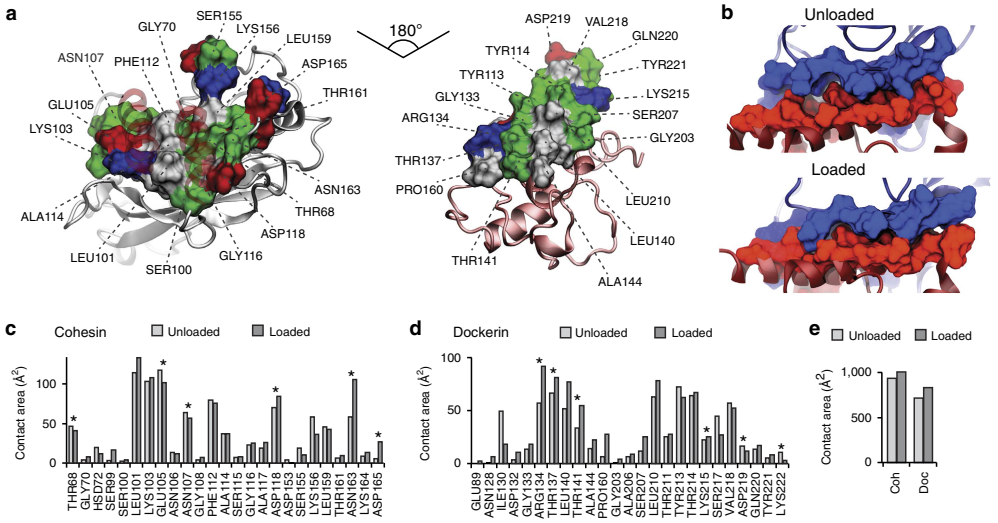


Figure 3 | Analysis of binding interface and catch bond mechanism from SMD. (**a**) Surface plots for the main interacting residues of Coh (left) and Doc (right). Hydrophobic residues are shown in grey, polar residues in green, and negative and positive residues in red and blue, respectively. Both Coh and Doc exhibit a hydrophobic patch in the centre of the binding surface that is surrounded by polar and charged residues. (**b**) Rearrangement of binding residues of Coh (blue) and Doc (red) under force. Following mechanical loading, an interdigitated complex is formed that resembles teeth of a zipper. (**c**,**d**) Surface contact area of interacting residues of Coh (c) and Doc (d) in the absence and presence of force. Residues forming prevalent hydrogen bonds are indicated with stars. (**e**) Total contact surface area of Coh and Doc in unloaded and loaded conformations.

rate dependence for this unfolding event is shown in Supplementary Fig. 3.

The exceptionally high rupture forces measured experimentally (Fig. 2) are hugely disproportionate to the XMod-Doc:Coh biochemical affinity, which at $K_D \sim 20$ nM (ref. 12) is comparable to typical antibody–antigen interactions. Antibody–antigen interactions, however, will rupture at only ~ 60 pN at similar loading rates[24], while bimolecular complexes found in muscle exposed to mechanical loading *in vivo* will rupture at ~ 140 pN (ref. 25). Trimeric titin–telethonin complexes also found in muscle exhibit unfolding forces around 700 pN (ref. 26), while Ig domains from cardiac titin will unfold at ~ 200 pN (ref. 27). The XMod-Doc:Coh ruptures reported here fell in a range from 600 to 750 pN at loading rates ranging from 10 to 100 nN s^{-1}. At around half the rupture force of a covalent gold-thiol bond[28], these bimolecular protein rupture forces are, to the best of our knowledge, among the highest of their kind ever reported. The covalent bonds in this system are primarily peptide bonds in the proteins and C-C and C-O bonds in the PEG linkers. These are significantly more mechanically stable than the quoted gold-thiol bond rupture force (~ 1.2 nN) (ref. 29) and fall in a rupture force range > 2.5 nN at similar loading rates. Therefore, breakage of covalent linkages under our experimental conditions is highly unlikely. We note that the high mechanostability observed here is not the result of fusing the proteins to the CBM or Xyn domains. The covalent linkages and pulling geometry are consistent with the wild-type complex and its dissociation pathway. *In vivo*, the Coh is anchored to the peptidoglycan cell wall through its C-terminal sortase motif. The XMod–Doc is attached to the cellulose substrate through two N-terminal CBM domains. By pulling the XMod–Doc through an N-terminal Xyn fusion domain, and the Coh through a C-terminal CBM, we established an experimental pulling geometry that matches loading of the complex *in vivo*. This pulling geometry was also used in all simulations. The discontinuity between its commonplace biochemical affinity and remarkable resistance to applied force illustrates how this complex is primed for mechanical stability and highlights differences in the unbinding pathway between dissociation at equilibrium and dissociation induced mechanically along a defined pulling coordinate.

Steered molecular dynamics. To elucidate the molecular mechanisms at play that enable this extreme mechanostability, we carried out all-atom steered molecular dynamics (SMD) simulations. The Xyn and CBM domains were not modelled to keep the simulated system small and reduce the usage of computational resources. This approximation was reasonable as we have no indication that these domains significantly affect the XMod–Doc:Coh binding strength[30]. After equilibrating the crystal structure[12], the N-terminus of XMod–Doc was harmonically restrained while the C-terminus of Coh was pulled away at constant speed. The force applied to the harmonic pulling spring was stored at each time step. We tested pulling speeds of 0.25, 0.625 and 1.25 Å ns^{-1}, and note that the slowest simulated pulling speed was $\sim$4,000 times faster than our fastest experimental pulling speed of 6.4 µm s^{-1}. This difference is considered not to affect the force profile, but it is known to account for the scale difference in force measured by SMD and AFM[31,32].

SMD results showed the force increased with distance until the complex ruptured for all simulations. At the slowest pulling speed of 0.25 Å ns^{-1} the rupture occurred at a peak force of ~ 900 pN, as shown in Supplementary Fig. 4 and Supplementary Movie 1. We analysed the progression and prevalence of hydrogen bonded contacts between the XMod–Doc and Coh domains to identify

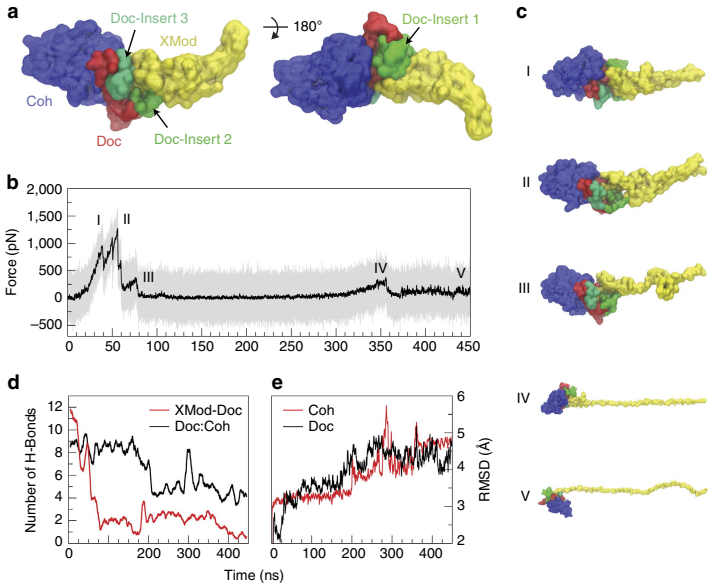


Figure 4 | SMD shows unfolding of XMod destabilizes Doc:Coh binding interface. XMod was unfolded by moving the harmonic restraint to the C terminus of XMod while the N terminus was moved at 0.625 Å ns^{-1}. (**a**) Surface representation of XMod-Doc:Coh complex with Doc insert sequences. Coh is shown in blue, Doc in red and green (inserts), and XMod in yellow. (**b**) Force time trace of XMod unfolding. The domain starts to unfold in several substeps starting at ~400 pN. Snapshots at different time steps are labelled I-V and are shown in (**c**). Steps IV and V are shown at smaller scale. (**d**) Average number of hydrogen bonds between Doc:Coh (black) and XMod-Doc (red). XMod-Doc contact is dominated by the insert sequences 1–3. (**e**) Root mean squared deviation (RMSD) of Doc (black) and Coh (red).

key residues in contact throughout the entire rupture process and particularly immediately before rupture. These residues are presented in Fig. 3a,c,d and Supplementary Figs 5,6. The simulation results clearly reproduced key hydrogen bonding contacts previously identified[12] as important for Doc:Coh recognition (Supplementary Fig. 5).

The main interacting residues are shown in Fig. 3a,b. Both Coh and Doc exhibit a binding interface consisting of a hydrophobic centre (grey) surrounded by a ring of polar (green) and charged residues (blue, positive; red, negative). This residue pattern suggests the hydrophilic side chains protect the interior hydrophobic core from attack by water molecules, compensating for the flat binding interface that lacks a deep pocket. The geometry suggests a penalty to unbinding that stabilizes the bound state. Further, we analysed the contact surface areas of interacting residues (Fig. 3b–e). The total contact area was found to increase due to rearrangement of the interacting residues when the complex is mechanically stressed, as shown in Fig. 3e and Supplementary Movie 2. Doc residues in the simulated binding interface clamped down on Coh residues upon mechanical loading, resulting in increased stability and decreased accessibility of water into the hydrophobic core of the bound complex (Fig. 3b). These results suggest that a catch bond mechanism is responsible for the remarkable stability[33] under force and provide a molecular mechanism which the XMod–Doc:Coh complex uses to summon mechanical strength when needed, while still allowing relatively fast assembly and disassembly of the complex at equilibrium. The residues that increase most in contact area (Fig. 3c,d) present promising candidates for future mutagenesis studies.

Among the 223 Doc sequences from *R. flavefaciens*, six subfamilies have been explicitly identified using bioinformatics approaches[8]. The XMod–Doc investigated here belongs to the 40-member Doc family 4a. A conserved feature of these Doc modules is the presence of three sequence inserts that interrupt the conserved duplicated F-hand motif Doc structure. In our system, these Doc sequence inserts make direct contacts with XMod in the crystallized complex (Fig. 1) and suggest an interaction between XMod and Doc that could potentially propagate to the Doc:Coh binding interface. To test this, an independent simulation was performed to unfold XMod (Fig. 4). The harmonic restraint was moved to the C-terminus of XMod so that force was applied from the N- to C-terminus of XMod only, while leaving Doc and Coh unrestrained. The results (Fig. 4b) showed XMod unfolded at forces slightly higher than but similar to the XMod–Doc:Coh complex rupture force determined from the standard simulation at the same pulling speed. This suggested XMod unfolding before Doc:Coh rupture was not probable, but could be observed on occasion due to the stochastic nature of domain unfolding. This was consistent with experiments where XMod unfolding was observed in ~35–40% of traces. Furthermore, analysis of the H-bonding between Doc and XMod (Fig. 4d, red) indicated loss of contact as XMod unfolded, dominated by contact loss between the three Doc insert sequences and XMod. Interestingly, XMod unfolding clearly led to a decrease in H-bonding between Doc and Coh at a later stage (~200 ns) well after XMod had lost most of its contact with Doc, even though no force was being applied across the Doc:Coh binding interface. This provided evidence for direct stabilization of the Doc:Coh binding interface by XMod.

As shown in Fig. 4e, the root mean squared deviation (RMSD) of Doc increased throughout the simulation as XMod unfolded. Coh RMSD remained stable until it started to lose H-bonds with Doc. Taken together this suggests that, as XMod unfolded, Coh and Doc became more mobile and lost interaction strength, potentially explaining the increase in Δx from 0.13 to 0.19 nm on unfolding of XMod in the experimental data sets. Apparently the XMod is able to directly stabilize the Doc:Coh interface, presumably through contact with Doc insert sequences that then propagate this stabilizing effect to the Doc:Coh binding interface.

In summary, we investigated an ultrastable XMod-Doc:Coh complex involved in bacterial adhesion to cellulose. While previously the role of XMod functioning in tandem XMod-Doc dyads was unclear[12,14], we show that XMod serves as a mechanical stabilizer and force-shielding effector subdomain in the ultrastable ligand–receptor complex. The Doc:Coh complex presented here exhibits one of the most mechanically robust protein–protein interactions reported thus far, and points towards new mechanically stable artificial multi-component biocatalysts for industrial applications, including production of second-generation biofuels.

Methods

Site-directed mutagenesis. Site-directed mutagenesis of *R. flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus*[12] were subjected to QuikChange mutagenesis[34] to install the following mutations: A2C in the CBM and T129C in the xylanase, respectively.

For the construction of the native configuration of the CohE-CBM A2C fusion protein Gibson assembly[35] was used. For further analysis CohE-CBM A2C was modified with a QuikChange PCR[36] to replace the two cysteins (C2 and C63) in the protein with alanine and serine (C2A and C63S). All mutagenesis products were confirmed by DNA sequencing analysis.

The XynT6-XDoc T129C was constructed using the following primers:
5′-acaaggaaggtaagccaatggttaatgaatgcgatccagtgaaacgtgaac-3′
5′-gttcacgtttcactggatcgcattcattaaccattggcttaccttccttgt-3′

The CBM-CohE A2C was constructed using the following primers:
5′-ttaactttaagaaggagatataccatgtgcaatacaccggtatcaggcaatttgaag-3′
5′-cttcaaattgcctgataccggtgtattgcacatggtatatctccttcttaaagttaa-3′

The CohE-CBM C2A C63S was constructed using the following phosphorylated primers:
5′-ccgaatgccatggccaatacaccgg-3′
5′-cagaccttctggagtgaccatgctgc-3′

Expression and purification of Xyn-XMod-Doc. The T129C Xyn-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with tris-buffered saline (TBS) buffer containing 20 mM imidazole and 2 mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialysed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at − 20 °C. The concentrations of the protein stock solutions were determined to be ∼5 mg ml^{-1} by absorption spectrophotometry.

Expression and purification of Coh-CBM. The Coh-CBM C2A, C63S fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicol containing ZYM-5052 media[37] overnight at 22 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at − 80 °C. The concentrations of the protein stock solutions were determined to be ∼5 mg ml^{-1} by absorption spectrophotometry.

Sample preparation. In sample preparation and single-molecule measurements calcium supplemented TBS buffer (Ca-TBS) was used (25 mM TRIS, 72 mM NaCl, 1 mM $CaCl_2$, pH 7.2). Cantilevers and cover glasses were functionalized according to previously published protocols[18,38]. In brief, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyn-XMod-Doc proteins were reduced for 2 h at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with Ca-TBS measurement buffer, centrifuged at 850 r.c.f. for 3 min, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with Ca-TBS to remove unbound proteins and stored under Ca-TBS before force spectroscopy measurements. Site-specific immobilization of the Coh-CBM-ybbR fusion proteins to previously PEGylated cantilevers or coverglasses was carried out according to previously published protocols[39]. In brief, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) for 1 h at room temperature. Levers or surfaces were then rinsed with Ca-TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 h at room 37°. Finally, surfaces or levers were subjected to a final rinse with Ca-TBS and stored under Ca-TBS before measurement.

Single-molecule force spectroscopy measurements. SMFS measurements were performed on a custom built AFM[40] controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise/equipartition method[41]. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2 to 6.4 μm s^{-1}. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18–24 h. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw data sets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single-molecule interactions. We filtered the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: Xyn (∼89 nm) and CBM (∼56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

Data analysis. Data were analysed using previously published protocols[17,18,22]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11$ nm connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. Detailed description of the contour length transformation can be found in Supplementary Note 1 and Supplementary Fig. 1.

For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs time trace in the immediate vicinity before the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying Gaussian fits to histograms of rupture forces and loading rates at each pulling speed.

Molecular dynamics simulations. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97 Å resolution and is available at the protein data bank (PDB:4IU3). A protonation analysis performed in VMD[42] did not suggest any extra protonation and all the amino-acid residues were simulated with standard protonation states. The system was then solvated, keeping also the water molecules present in the crystal structure, and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter ions, which were randomly arranged in the solvent. Two other systems, based on the aforementioned one, were created using a similar salt concentration to the one used in the experiments (75 mM of NaCl). This additional salt caused little or no change in SMD results. The overall number of atoms included in MD simulations varied from 300,000 in the majority of the simulations to 580,000 for the unfolding of the X-Mod.

The MD simulations in the present study were performed employing the NAMD molecular dynamics package[43,44]. The CHARMM36 force field[45,46] along with the TIP3 water model[47] was used to describe all systems. The simulations were done assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were

treated using the particle-mesh Ewald (PME)[48] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[44] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations all the systems were submitted to an energy minimization protocol for 1,000 steps. The first two nanoseconds of the simulations served to equilibrate systems before the production runs that varied from 40 to 450 ns in the 10 different simulations that were carried out. The equilibration step consisted of 500 ps of simulation where the protein backbone was restrained and 1.5 ns where the system was completely free and no restriction or force was applied. During the equilibration the initial temperature was set to zero and was constantly increased by 1 K every 100 MD steps until the desired temperature (300 K) was reached.

To characterize the coupling between Doc and Coh, we performed SMD simulations[49] of constant velocity stretching (SMD-CV protocol) employing three different pulling speeds: 1.25, 0.625 and 0.25 Å ns^{-1}. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically (center of mass of ASN5), and moving a second restraint point, at the end of the Coh domain (center of mass of GLY210), with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the *z*-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the *z*-axis. Owing to the flexibility of the linkers, this approach reproduces the experimental set-up. All analyses of MD trajectories were carried out employing VMD[42] and its plug-ins. Secondary structures were assigned using the Timeline plug-in, which employs STRIDE criteria[50]. Hydrogen bonds were assigned based on two geometric criteria for every trajectory frame saved: first, distances between acceptor and hydrogen should be <3.5 Å; second, the angle between hydrogen-donor-acceptor should be $<30°$. Surface contact areas of interacting residues were calculated employing Volarea[51] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling spherical probe that is screened around the area of Doc exposed to Coh and also Coh area exposed to Doc.

References

1. Doi, R. H. & Kosugi, A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* **2,** 541–551 (2004).
2. Carvalho, A. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin—dockerin complex. *Proc. Natl Acad. Sci. USA* **100,** 13809–13814 (2003).
3. Smith, S. P. & Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* **23,** 686–694 (2013).
4. Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. From cellulosomes to cellulosomics. *Chem. Rec.* **8,** 364–377 (2008).
5. Demain, A. L., Newcomb, M. & Wu, J. H. D. Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* **69,** 124–154 (2005).
6. Jindou, S. *et al.* Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. *FEMS Microbiol. Lett.* **285,** 188–194 (2008).
7. Ding, S. Y. *et al.* Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* **183,** 1945–1953 (2001).
8. Rincon, M. T. *et al.* Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS ONE* **5,** e12476 (2010).
9. Himmel, M. E. *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315,** 804–807 (2007).
10. Fierobe, H.-P. *et al.* Degradation of cellulose substrates by cellulosome chimeras Substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* **277,** 49621–49630 (2002).
11. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl Acad. Sci. USA* **106,** 13791–13796 (2009).
12. Salama-Alber, O. *et al.* Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288,** 16827–16838 (2013).
13. Adams, J. J., Webb, B. A., Spencer, H. L. & Smith, S. P. Structural characterization of type ii dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. *Biochemistry* **44,** 2173–2182 (2005).
14. Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **103,** 305–310 (2006).
15. Sikora, M. & Cieplak, M. Mechanical stability of multidomain proteins and novel mechanical clamps. *Proteins Struct. Funct. Bioinf.* **79,** 1786–1799 (2011).
16. Brunecky, R. *et al.* Structure and function of the Clostridium thermocellum cellobiohydrolase A X1-module repeat: enhancement through stabilization of the CbhA complex. *Acta. Crystallogr.* **68,** 292–299 (2012).
17. Stahl, S. W. *et al.* Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **109,** 20431–20436 (2012).
18. Jobst, M. A., Schoeler, C., Malinowska, K. & Nash, M. A. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.* **82,** e50950 (2013).
19. Otten, M. *et al.* From genes to protein mechanics on a chip. *Nat. Methods* **11,** 1127–1130 (2014).
20. Livadaru, L., Netz, R. R. & Kreuzer, H. J. Stretching response of discrete semiflexible polymers. *Macromolecules* **36,** 3732–3744 (2003).
21. Hugel, T., Rief, M., Seitz, M., Gaub, H. & Netz, R. Highly stretched single polymers: atomic-force-microscope experiments versus *ab-initio* theory. *Phys. Rev. Lett.* **94,** 048301 (2005).
22. Puchner, E. M., Franzen, G., Gautel, M. & Gaub, H. E. Comparing proteins by their unfolding pattern. *Biophys. J.* **95,** 426–434 (2008).
23. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor—ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).
24. Morfill, J. *et al.* Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. *Biophys. J.* **93,** 3583–3590 (2007).
25. Berkemeier, F. *et al.* Fast-folding α-helices as reversible strain absorbers in the muscle protein myomesin. *Proc. Natl Acad. Sci. USA* **108,** 14139–14144 (2011).
26. Bertz, M., Wilmanns, M. & Rief, M. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc. Natl Acad. Sci. USA* **106,** 13307–13310 (2009).
27. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402,** 100–103 (1999).
28. Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond? *Science* **283,** 1727–1730 (1999).
29. Xue, Y., Li, X., Li, H. & Zhang, W. Quantifying thiol-gold interactions towards the efficient strength control. *Nat. Commun.* **5,** 4348 (2014).
30. Bomble, Y. J. *et al.* Modeling the self-assembly of the cellulosome enzyme complex. *J. Biol. Chem.* **286,** 5614–5623 (2011).
31. Sotomayor, M. & Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **316,** 1144–1148 (2007).
32. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **271,** 997–999 (1996).
33. Thomas, W. *et al.* Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophys. J.* **90,** 753–764 (2006).
34. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques* **26,** 680–682 (1999).
35. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
36. Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28,** e78 (2000).
37. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.* **41,** 207–234 (2005).
38. Zimmermann, J. L., Nicolaus, T., Neuert, G. & Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat. Protoc.* **5,** 975–985 (2010).
39. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1,** 280–285 (2006).
40. Gumpp, H., Stahl, S. W., Strackharn, M., Puchner, E. M. & Gaub, H. E. Ultrastable combined atomic force and total internal fluorescence microscope. *Rev. Sci. Instrum.* **80,** 063704 (2009).
41. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868 (1993).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14,** 33–38 (1996).
43. Kalé, L. *et al.* NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151,** 283–312 (1999).
44. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
45. Best, R. B. *et al.* Optimization of the additive CHARMM All-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ 1and χ 2dihedral Angles. *J. Chem. Theory Comput.* **8,** 3257–3273 (2012).
46. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
47. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–934 (1983).
48. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

49. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72,** 1568–1581 (1997).
50. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinf.* **23,** 566–579 (1995).
51. Ribeiro, J. V., Tamames, J. A. C., Cerqueira, N. M. F. S. A., Fernandes, P. A. & Ramos, M. J. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug Des.* **82,** 743–755 (2013).

Acknowledgements

We gratefully acknowledge funding from an advanced grant of the European Research Council (Cellufuel Grant 294438) and from DFG SFB 1032 and the Excellence Cluster Center for Integrated Protein Science Munich. This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and the National Science Foundation (NSF, MCB-1157615 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and the NCSA Blue Waters sustained-petascale supercomputer as part of the general allocations (Simulations of Cellulosomal Subunits: Components of a Molecular Machinery for Depolymerization of Feedstock for Production of Second Generation Biofuels, to K.S.). A grant to E.A.B., H.E.G. and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science - The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland.

Author contributions

Performed and analysed SMFS experiments: C.S., K.H.M., L.F.M., M.A.J., E.D. and M.A.N.; performed and analysed MD simulations: R.C.B. and K.S.; provided proteins and DNA cloning vectors: W.O., D.B.F. and E.A.B.; wrote and edited the manuscript: C.S., K.H.M., R.C.B., E.A.B., K.S., H.E.G. and M.A.N.; supervised research: E.A.B., K.S., H.E.G. and M.A.N.

Additional information

Accession codes: Plasmids used in this study are available through Addgene (https://www.addgene.org) under the following accession codes: Xylanase-Xmodule-Dockerin: 60865; Cohesin-CBM: 60866.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* 5:5635 doi: 10.1038/ncomms6635 (2014).

Appendix C

Letter in Nano Letters
Chapter 4 in original form

Mapping Mechanical Force Propagation Through Biomolecular Complexes

by

Constantin Schoeler, Rafael C. Bernardi, Klara H. Malinowska, Ellis Durner, Wolfgang Ott, Edward A. Bayer, Klaus Schulten, Michael A. Nash, and Hermann E. Gaub

published in

Nano Letters 15, 7370-7376 doi:10.1021/acs.nanolett.5b02727 (2015)
Reprinted from Schoeler *et al.* [72] with permission from the American Chemical Society

NANO LETTERS
Letter
pubs.acs.org/NanoLett

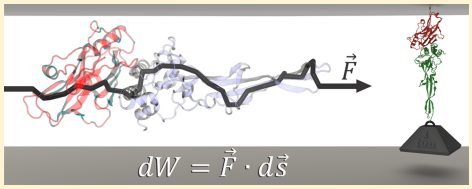
Mapping Mechanical Force Propagation through Biomolecular Complexes

Constantin Schoeler,[†,#] Rafael C. Bernardi,[‡,#] Klara H. Malinowska,[†] Ellis Durner,[†] Wolfgang Ott,[†,§] Edward A. Bayer,[∥] Klaus Schulten,[‡,⊥] Michael A. Nash,[*,†] and Hermann E. Gaub[†]

[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[‡]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[§]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany
[∥]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
[⊥]Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Here we employ single-molecule force spectroscopy with an atomic force microscope (AFM) and steered molecular dynamics (SMD) simulations to reveal force propagation pathways through a mechanically ultrastable multidomain cellulosome protein complex. We demonstrate a new combination of network-based correlation analysis supported by AFM directional pulling experiments, which allowed us to visualize stiff paths through the protein complex along which force is transmitted. The results implicate specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces.



KEYWORDS: *Force propagation, single molecule force spectroscopy, steered molecular dynamics, network analysis, cohesin−dockerin*

Mechanical forces play a fundamental role in biological systems. Cells are able to sense and respond to mechanical cues in their environment by, for example, modulating gene expression patterns,[1] reshaping the extracellular matrix,[2] or exhibiting differential biochemical activities.[3] At the molecular level, these behaviors are governed by mechanically active proteins. Such proteins are able to sense and respond to force by undergoing conformational changes,[4] exposing cryptic binding sequences,[5] acting synergistically with ion channels,[6] or modulating their function in a variety of ways.[7−9]

Experimental methods including AFM single-molecule force spectroscopy (SMFS) allow direct measurement of molecular mechanical properties. These studies have demonstrated the importance of the shear topology involving parallel breakage of hydrogen bonds in providing mechanical stability to protein folds.[10,11] Many globular domains and protein complexes also exhibit a directional dependence in unfolding mechanics, consisting of stiff and soft axes.[12−18] Pulling geometry can be defined by controlling the positions of the chemical linkages between protein monomer units through a variety of bioconjugate techniques.

Primary sequences of mechanically active proteins are extremely diverse, essentially rendering them undetectable by conventional bioinformatics approaches. Yet, another computational approach, namely, molecular dynamics (MD), allows sampling of structural conformations of large and frequently mechanostable protein complexes.[19,20] Analysis of these conformations from MD trajectories have recently led to the development of network-based correlation methods for investigating signal transmission and allosteric regulation in proteins.[21−23] In network models, local correlations of positional fluctuations in a protein are represented as a web of inter-residue connections. Within such a network, the behavior of nodes that are highly correlated and within close physical proximity can be analyzed to obtain the shortest path between two network nodes (i.e., amino acids). This analysis helps to identify which connecting residues are most important for intramolecular communication.[23−25] Examination of multiple pathways, also known as suboptimal paths, within an acceptable deviation from the optimal path helps to detect the web of nodes critical for transmission of information.

Among MD methods, steered molecular dynamics (SMD) simulations in which external forces are used to explore the response and function of proteins have become a powerful tool especially when combined with SMFS.[6] SMD has been successfully employed in a wide range of biological systems, from the investigation of protein mechanotransduction,[5,26] to

Received: July 9, 2015
Revised: August 5, 2015
Published: August 11, 2015

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

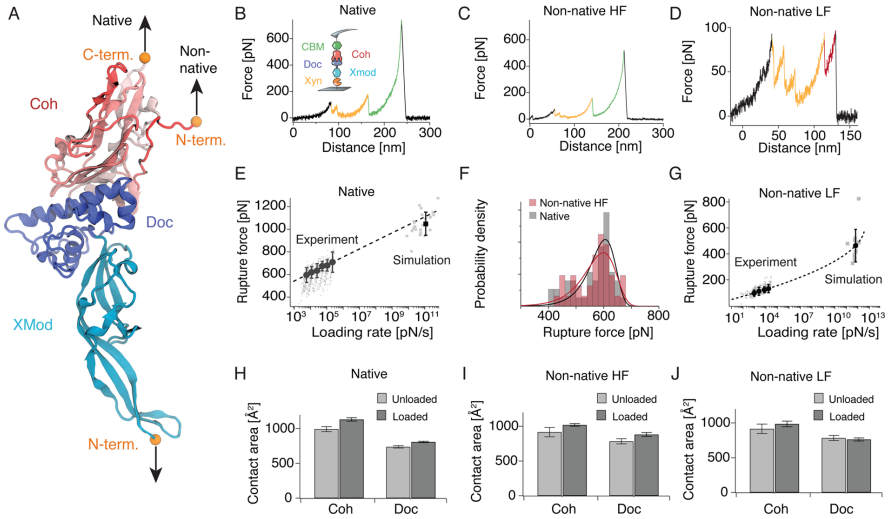


Figure 1. Single molecule force spectroscopy and steered molecular dynamics of XMod-Doc:Coh in two pulling configurations. (A) Crystal structure of the XMod-Doc:Coh complex (PDB 4IU3) with orange spheres marking the termini where force was applied. (B) Experimental unfolding trace for the native pulling configuration at a pulling speed of 1600 nm s^{-1}. The inset shows a schematic of the pulling geometry. Unfolding signatures of the Xyn and CBM marker domains are marked in orange and green, respectively. (C) Experimental unfolding trace for the non-native high force class obtained at a pulling speed of 700 nm s^{-1}. (D) Experimental unfolding trace for the non-native low force class obtained at a pulling speed of 700 nm s^{-1}. The additional 17−19 nm contour length increment attributed to N-terminal Coh unfolding is shown in red. (E) Dynamic force spectrum for XMod-Doc:Coh unbinding in the native geometry obtained from experiment and simulations. Gray points and squares represent the rupture force/loading rate pairs obtained from experiment and simulation, respectively. Black circles represent the most probable rupture force/loading rate obtained from Gaussian fits to the experimental data at six pulling speeds. The black square shows the mean rupture force and loading rate for the simulated rupture events. (F) Rupture force histograms obtained at a pulling speed of 800 nm s^{-1} for the native (gray, n = 46) and non-native high force class (red, n = 48). Fitted probability densities $p(F)$ are shown as solid black and red lines. Data for both pulling configurations were obtained with the same cantilever to minimize calibration errors. (G) Dynamic force spectrum for XMod-Doc:Coh unbinding in the non-native low force class obtained from experiments and simulation. The same representation as in (E) is used. (H,I,J) Unloaded and loaded surface contact areas for the different pulling geometries ((H) native, (I) non-native high force class, and (J) non-native low force class).

permeability of membrane channels,[27,28] and the characterization of protein−receptor interactions.[29] SMD simulations have also been used to study force propagation through proteins by employing force distribution analysis (FDA).[30,31] In FDA, all pairwise forces, which are usually calculated in MD simulations, are stored in $N \times N$ matrices, where N is the number of atoms.[32] These pairwise forces can then be used to assess a protein's response to a mechanical or allosteric signal.[33] In the FDA approach, atoms under mechanical strain are identified by subtracting forces of both loaded and unloaded states for each pair of interacting atoms.[31] However, to achieve a sufficient signal-to-noise ratio, FDA will often require exhaustive sampling of the conformational space.[32,34] FDA, therefore, requires more computational resources than usual SMD studies, which are frequently already computationally demanding. There is therefore a clear need for new analysis methods that enable visualization of force propagation pathways from a single SMD trajectory.

Here we implemented a novel combination of SMD, network-based correlation analysis, and thermodynamic fluctuation theory, supported by AFM-SMFS experiments to study force propagation through a protein complex subjected to different pulling geometries. We chose an ultrastable receptor−ligand interaction as a model system because of its remarkably high mechanical stability,[29] which effectively improves the signal-to-noise ratio. This complex consists of two interacting protein domains called cohesin (Coh) and dockerin (Doc) that maintain bacterial adhesion of *Ruminococcus flavefaciens* to cellulosic substrates. Doc is found within the same polypeptide chain as a stabilizing ancillary domain called X-module (XMod), located N-terminally of Doc. Based on its position with the *R. flavefaciens* cellulosomal network, Coh is mechanically anchored *in vivo* at its C-terminal end to the cell surface. Our prior work demonstrated that, when force is applied to the complex in the native configuration (i.e., C-terminal Coh, N-terminal XMod-Doc anchor points), the complex is extremely stable, exhibiting high rupture forces of 600−750 pN at loading rates from 1−100 nN s^{-1}.[29] Since the bulk equilibrium affinity of the complex is an unremarkable 20 nM,[35] we hypothesized that the high mechanostability is explained by a catch bond mechanism. AFM rupture force data and SMD simulations supported this prediction, where it was observed that the contact surface area of the two proteins increased as mechanical force was applied.

To characterize the mechanisms behind Coh:Doc high stability, here we additionally pulled the complex apart in a non-native configuration (i.e., N-terminal Coh, N-terminal XMod-Doc anchor points). In the non-native pulling

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

configuration, we found that the complex dissociated along two competing pathways with very different mechanical characteristics.

Our new dynamic network analysis protocol reveals how different mechanical behaviors are attributable to differences in the direction of force transmission across the binding interface. Together, the experiments and simulations depict a simple physical mechanism for achieving high complex rupture forces: the complex directs force along pathways orthogonal to the pulling axis.

Single-Molecule Pulling Experiments and SMD. For SMFS experiments, XMod-Doc was produced as a fusion protein with an N-terminal Xylanase (Xyn) domain. Coh was produced as either an N- or C-terminal fusion domain with a carbohydrate binding module (CBM). These fusion domains were used for site specific immobilization to a glass surface and AFM cantilever to achieve the two loading configurations shown in Figure 1A and further served as marker domains with known unfolding length increments to validate single-molecule interactions and sort SMFS data traces.[36]

For the native pulling configuration found *in vivo*, Coh-CBM and XMod-Doc are loaded from their C- and N-termini, respectively (Figure 1A). A representative unbinding trace for the native pulling configuration is shown in Figure 1B. We measured the loading rate dependence of complex rupture using both experimental and SMD data sets (unbinding trace from SMD shown in Figure 3A) and plotted them on a combined dynamic force spectrum (Figure 1E). The linear Bell model produced fit parameters for the effective distance to the transition state Δx = 0.13 nm, and the zero-force off rate k_{off} = 4.7×10^{-4} s^{-1}. Both experimental and simulation data are well described by a single Bell expression, despite the differences in loading rates between experiments and simulation. The observation suggests that the application of force does not significantly change Δx for this particular configuration.

To test the influence of pulling geometry on mechanical stability, we performed SMFS and SMD on the system where Coh was pulled from the opposite terminus (i.e., non-native N-terminus, cf. Figure 1A). Unlike the native pulling geometry, this geometry exhibited two clearly distinct unbinding pathways that are characterized by different force ranges (high or low) at which the complex dissociated. We refer to these pathways as non-native high force (HF) (Figure 1C) and non-native low force (LF) (Figure 1D).

AFM data traces classified as non-native HF showed similar characteristics as those in the native pulling configuration (cf. Figure 1B,C,F). The non-native LF traces, however, exhibited a markedly different unfolding behavior (Figure 1D). Xyn unfolding (highlighted in orange) was regularly observed, but CBM unfolding was only very rarely observed. The complex usually did not withstand forces high enough to unfold CBM when rupturing along the non-native LF path. Among non-native LF curves, we regularly found an additional contour length increment of 17−19 nm consistent with unfolding of ∼60 amino acids located at the N-terminus of Coh. This unfolding occurred immediately following Xyn unfolding (Figure 1D, red), or alternatively prior to Xyn unfolding, or with a substep (Supplementary Figure S1). Taken together, it appears that partial Coh unfolding from the N-terminus destabilizes the complex, causing lower rupture forces (Figure 1G).

The experimental rupture forces from the non-native HF class were indistinguishable from those arising in the native configuration. To confirm this, we performed additional measurements where both Coh configurations were alternately probed with the same Xyn-XMod-Doc functionalized cantilever (Supplementary Figure S2), eliminating inaccuracies introduced through multiple cantilever calibration. Most probable rupture forces at a pulling speed of 800 nm s^{-1} of 606 and 597 pN for the native configuration and non-native HF class, respectively, were determined in the Bell Evans model (Figure 1F, Supplementary eq S2), demonstrating that the native and non-native HF classes are experimentally indistinguishable.

For the LF class, we analyzed the final complex rupture event and plotted the combined dynamic force spectrum (Figure 1G). Here, simulated and experimentally observed data were not well described by a single Bell expression. In such cases nonlinear models have been developed to obtain kinetic and energetic information from dynamic force spectra.[37,38] To fit the combined data, we used the nonlinear Dudko−Hummer−Szabo (DHS) model (Supplementary eq S3) and obtained values of Δx = 0.42 nm and k_{off} = 0.005 s^{-1}. The DHS model further provides the free energy difference ΔG between the bound state and the transition state as a fit parameter, which was found to be ΔG = 129 k_BT. The model fit produced a distance to transition that was much longer than observed for the native configuration. Independent SMD simulations for the non-native pulling configuration were found to also lead to HF and LF unbinding scenarios (see below, Figure 4A,D, respectively).

The differential solvent contact area was calculated from SMD simulations to estimate the intermolecular contact area in the Doc:Coh complex. In the native configuration, the simulated Doc:Coh contact area increased by 14% and 9% for Coh and Doc, respectively (Figure 1H). For the non-native HF class, the contact area increased by 11% and 12% for Coh and Doc, respectively (Figure 1I). In the non-native LF class, the contact area increased by only 7% for Coh and decreased by 3% for Doc (Figure 1J). Evidently, an increased surface contact area for Doc in the native and non-native HF pathways correlated with high mechanostability of the system.

Force Propagation Theory: A Simple Model. To further understand the observed unbinding pathways, we sought to identify paths through the molecule along which the externally applied load propagates. From thermodynamic fluctuation theory,[39,40] it is known that the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i are related through

$$\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{1}$$

where $\Delta \mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$ and $\mathbf{r}_i$ is the position of atom i. The derivative on the right-hand side of eq 1 states that neighboring atoms i and j will move with high correlation due to an external force $\mathbf{F}_i$ acting on atom i if the coupling between them is strong. Hence, a given element of a correlation matrix $M_{ij} = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle$ will be large in the case of a strong interaction potential between i and j. When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation, while stiff degrees become more important for the dynamics of the system.

Consequently, paths with high correlation of motion describe the paths along which force propagates through the system. To illustrate this behavior for a toy system, we employed the NAMD[41] SMD[42] constant velocity protocol to a test pattern of identical spheres connected with harmonic springs of different

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

stiffness (Figure 2A). The position of one sphere was fixed during the simulation, while another sphere on the opposite

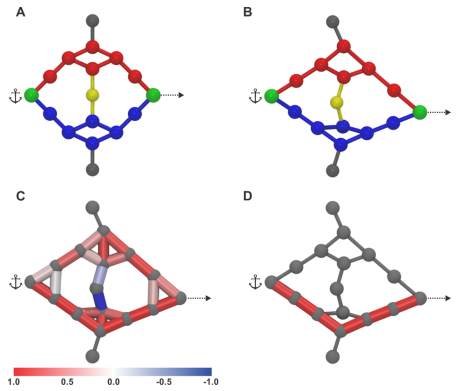


Figure 2. Network analysis test simulation. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red, k; blue, $5k$; yellow, $7.5k$; black, $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs were introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red.

side of the structure was withdrawn at constant velocity. The strained structure at the end of the simulation is shown in Figure 2B. We assigned weights to the lines between spheres according to the Pearson correlation coefficient C_{ij} (Supplementary eq S4) between those network nodes (Figure 2C). The Pearson correlation coefficient differs from the left-hand side of eq 1 by a normalization factor $(\langle \Delta \mathbf{r}_i^2(t)\rangle \langle \Delta \mathbf{r}_j^2(t)\rangle)^{-1/2}$ and was chosen to make our analysis mathematically more tractable. For a detailed discussion on this choice of correlation measure, see Supporting Information. In a harmonic potential approximation, the equipartition theorem can be applied to this normalization factor resulting in the following expression for C_{ij}:

$$C_{ij} = \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \sqrt{k_{i,\mathrm{eff}} k_{j,\mathrm{eff}}} \tag{2}$$

where $k_{i,\mathrm{eff}} = \left(\frac{1}{k_{x_i}} + \frac{1}{k_{y_i}} + \frac{1}{k_{z_i}}\right)^{-1}$ and k_{x_i} is the curvature of the potential on atom i in the x direction. For a full derivation, see Supporting Information. Equation 2 illustrates how Pearson correlation is a suitable measure to identify the stiff paths in our simple model. We then used dynamical network analysis implemented in VMD[49] to find the path of highest correlation (Figure 2D). As expected from eq 1, we found this path to be the one connected by the stiff springs.

Force Propagation through XMod-Doc:Coh Complex. The simple pattern of spheres validated our general approach of using local correlations to identify load-bearing pathways through networks. We next employed dynamical network analysis to understand force propagation through the XMod-Doc:Coh complex.

The dynamic networks for the native configuration (unloaded and loaded) are shown in Figure 3B,C, respectively. While the network shows multiple suboptimal paths in the unloaded scenario, the loaded case exhibits a well-defined main path along which force propagates through the system. Interestingly, in the loaded configuration, force propagates through both binding helices of Doc, which results in a force path with large normal components to the unbinding axis close to the binding interface as illustrated in Figure 3D. It had been shown for another ultrastable protein, namely, silk crystalline units, that curving force paths distribute tension through the entire system.[31] A strategy that assumes an indirect path would therefore allow the system to have more time to absorb the tension from the applied force. The result here supports the view that directing the force along a path with significant perpendicular components to the pulling axis leads to high mechanical stability. In a simple mechanical picture, a certain amount of mechanical work, namely $dW = \mathbf{F}\cdot d\mathbf{s}$, is required to separate the two binding interfaces by a distance Δz and break the interaction. In this simplified picture, $d\mathbf{s}$ points along the unbinding axis, whereas the force $\mathbf{F}$ is locally largely perpendicular to this direction. Consequently, a larger force is required to break the interaction than in a scenario where the force path would point along the unbinding axis.

To validate this picture, we repeated the same analysis for the non-native HF and non-native LF pathways. The HF simulation (Figure 4A) exhibited only a small stretching of the flexible N-terminal region of Coh and complex dissociation at approximately 800 pN and a pulling distance around 10 nm. However, the LF case shown in Figure 4D exhibited a stepwise N-terminal Coh unfolding, dissociating at a force of about 480 pN at a pulling distance of about 25 nm. This behavior confirmed our assignment of the experimentally observed 17–19 nm contour length increment to Coh unfolding up to residue 62 in PDB 4IU3.

While the experimental data did not show a detectable difference between the native configuration and the non-native HF class, the propagation of force takes place along a different pathway (Figure 4B). For N-terminal Coh pulling, helix 3 of Doc is not involved in the propagation of force as it is for the native geometry. In the native configuration, force propagates through the center of Coh, while for non-native HF the path is shifted toward the side of the molecule. Despite these differences, there is a common feature between the native and non-native HF pathways. At the binding interface, the pathway again shows pronounced components perpendicular to the unbinding axis (cf. Figure 4C), suggesting that this feature is indeed responsible for the exceptional mechanical strength observed for these two unbinding pathways.

Figure 4E shows the force propagation pathway for the non-native LF class prior to rupture. Due to the unfolding of the N-terminal Coh segment, the propagation of force is shifted even further away from the central portion of Coh than for the non-native HF class. Interestingly, force is propagated through the small helical segment of Coh (ALA167-GLN179), a portion of the molecule that is not involved in force propagation for any of the other analyzed trajectories. Unlike in the aforementioned scenarios, there is no pronounced tendency for perpendicular force components at the binding interface for the non-native LF class. In fact, the force is propagated along a path largely parallel to the pulling axis (cf. Figure 4F). In cases where force

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370–7376

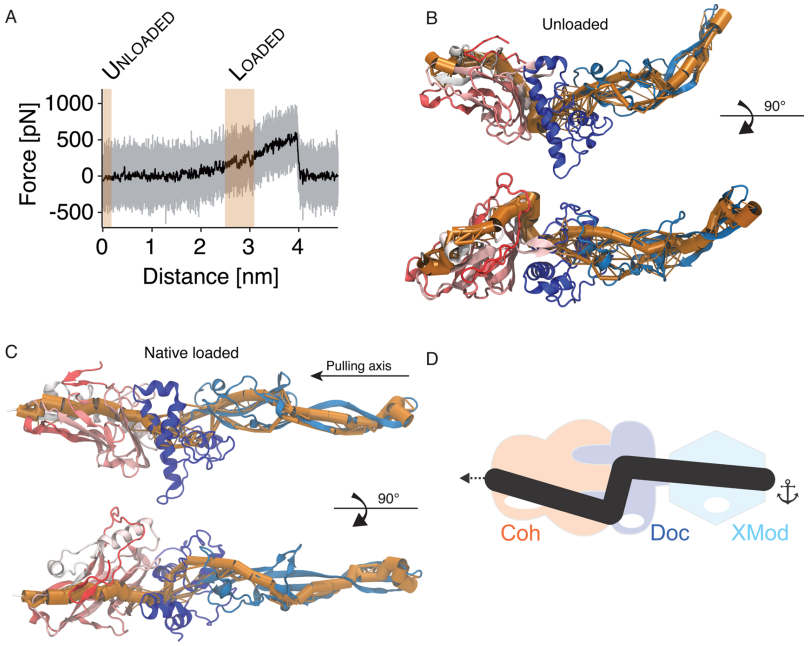


Figure 3. Force propagation through XMod-Doc:Coh in the native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray. The black line represents a moving average with a box size of 500 steps. The highlighted red areas denote the windows where dynamic networks and contact areas were calculated. (B) Network paths for the unloaded system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (C) Network paths for the loaded system. A detailed 2D representation of the pathway, highlighting the amino acids present in the pathway, is shown in Supplementary Figure S5. (D) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis.

propagation occurs parallel to the pulling axis, as in Figure 4E, low mechanical stability was observed.

The aforementioned force propagation architecture along with the effect of increasing contact surface area upon mechanical loading combine for elevated mechanostability of the system.[29] In cases where we observed an N-terminal Coh unfolding of 62 amino acids in the non-native geometry, the system was no longer able to summon this mechanism, causing dissociation at much lower forces.

Previously, our groups have reported on a family of mechanically stable protein ligand receptor complexes that are key building blocks of cellulosomes,[29,44−46] the multienzyme complexes used by select anaerobic bacteria to digest lignocellulose. However, the molecular origins of the stability of these complexes remained largely unclear. An initial clue was obtained when, in a previous work, we were able to show that contact surface area of the two proteins increased as mechanical force was applied.[29] In a different study,[47] coarse-grained MD simulations showed much smaller rupture forces at similar loading rates both for native and non-native pulling than we report here. This disagreement is likely due to the inability of the coarse-grained model to capture the rearrangement of amino acid side chains observed here. As we demonstrated, force propagation calculation from network-based correlation analysis helped in investigating the dramatic effect on the mechanical stability of the Doc:Coh interaction when different pulling geometries are applied. Our methodological approach, to the best of our knowledge, has never been applied even though network analysis of SMD trajectories was performed before to probe the mechanism of allosteric regulation in imidazole glycerol phosphate synthase.[48]

In summary, for both unbinding cases where we observed high mechanostability, we found that across the binding interface, force propagated along paths with strong normal components to the pulling direction. Such a behavior was not observed for the non-native LF class, where, presumably due to N-terminal Coh unfolding, the system was no longer able to direct the force across the binding interface at high angles. From these findings, we conclude that the ultrastable complex formed by Coh and Doc achieves its remarkable mechanostability by actively directing an externally applied force toward an unfavorable angle of attack at the binding interface, consequently requiring more force to achieve a given amount of separation along the pulling direction. Our results show that this mechanically stable complex uses an architecture that exploits simple geometrical and physical concepts from Newtonian mechanics to achieve high stability against external forces. The analytical framework derived here provides a basis for developing a deeper understanding of the functioning of various mechanoactive proteins that are crucial for physiolog-

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

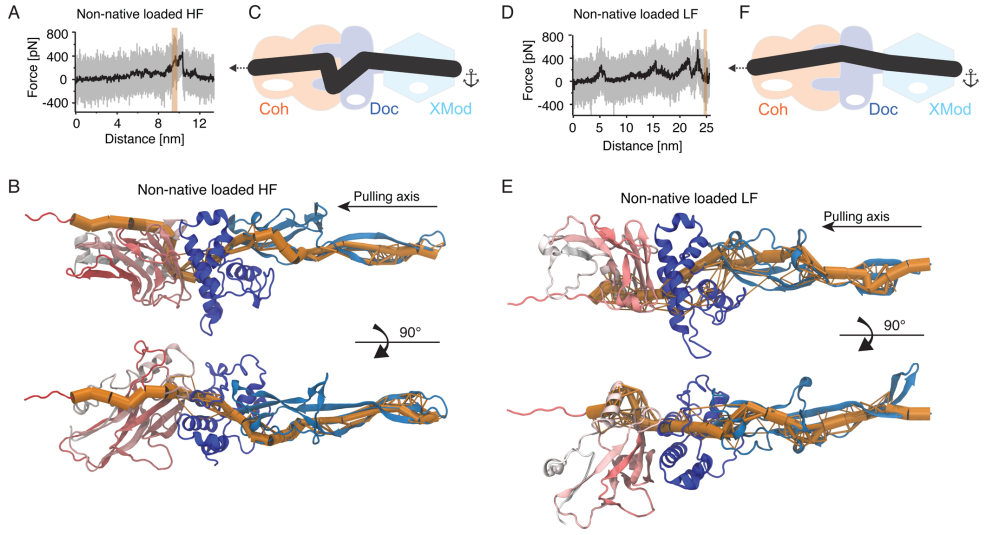


Figure 4. Force propagation through XMod-Doc:Coh in the non-native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500 steps. Note that this computational pulling experiment revealed a high-force behavior. (B) Network pathways calculated from dynamical network analysis for the non-native HF trajectory. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S6. (C) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis. (D) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500. This computational pulling experiment revealed partial Coh unfolding that led to LF behavior. (E) Network pathways for the non-native LF scenario. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S7. (F) Schematic model of force propagation across the Coh:Doc binding interface. Unlike in both HF scenarios, force propagates across the binding interface mostly along the unbinding axis.

ically relevant processes such as mechanotransduction, cellular mechanosensing, and pathogenesis. Additionally, it could provide a design platform for development of artificial mechanoactive systems with applications as tissue engineering scaffolds or components in engineered nanomaterials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b02727.

Details of experimental materials and methods, steered molecular dynamics simulations, and additional theoretical discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.nash@lmu.de.

Author Contributions

#These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Zaida Luthey-Schulten, Helmut Grubmüller, Marcelo Melo, and Jonathan Lai for helpful discussions. The authors gratefully acknowledge funding from an advanced grant to HEG from the European Research Council (Cellufuel Grant 294438). This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and from the Energy Biosciences Institute (UCB BP 2014OOO4J01 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and NERSC/Edison supercomputer as part of the DoE ALCC program. Resources of the National Energy Research Scientific Computing Center (NERSC) are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A grant to E.A.B., H.E.G., and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. The authors appreciate the support of the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nano-systems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science−The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland, and from the LMU Excellence Initiative.

REFERENCES

(1) Bao, G.; Suresh, S. *Nat. Mater.* **2003**, *2*, 715−725.
(2) Grodzinsky, A. J.; Levenston, M. E.; Jin, M.; Frank, E. H. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 691−713.
(3) Kumar, S.; Weaver, V. M. *Cancer Metastasis Rev.* **2009**, *28*, 113−127.
(4) Vogel, V.; Thomas, W. E.; Craig, D. W.; Krammer, A.; Baneyx, G. *Trends Biotechnol.* **2001**, *19*, 416−423.
(5) del Rio, A.; Perez-Jimenez, R.; Liu, R.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P. *Science* **2009**, *323*, 638−641.
(6) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−1148.
(7) Engel, A.; Muller, D. J. *Nat. Struct. Biol.* **2000**, *7*, 715−718.
(8) Noy, A.; Friddle, R. W. *Methods* **2013**, *60*, 142−150.
(9) De Yoreo, J. J.; Chung, S.; Friddle, R. W. *Adv. Funct. Mater.* **2013**, *23*, 2525−2538.
(10) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys. J.* **1998**, *75*, 3008−3014.
(11) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−1112.
(12) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. *Biophys. J.* **2012**, *103*, 2361−2368.
(13) Lee, W.; Zeng, X.; Rotolo, K.; Yang, M.; Schofield, C. J.; Bennett, V.; Yang, W.; Marszalek, P. E. *Biophys. J.* **2012**, *102*, 1118−1126.
(14) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820−17825.
(15) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12724−12728.
(16) Carrión-Vázquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738−743.
(17) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731−737.
(18) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(19) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64−74.
(20) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(21) Ribeiro, A. S. T.; Ortiz, V. *J. Phys. Chem. B* **2015**, *119*, 1835−1846.
(22) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.
(23) Van Wart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.
(24) Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z. *FEBS Lett.* **2010**, *584*, 376−386.
(25) Chennubhotla, C.; Bahar, I. *Mol. Syst. Biol.* **2006**, *2*, 1−13.
(26) Hytönen, V. P.; Vogel, V. *PLoS Comput. Biol.* **2008**, *4*, e24.
(27) Khalili-Araghi, F.; Gumbart, J.; Wen, P.-C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2009**, *19*, 128−37.
(28) Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 271−281.
(29) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; Gaub, H. E.; Nash, M. A. *Nat. Commun.* **2014**, *5*, 5635.
(30) Stacklies, W.; Vega, M. C.; Wilmanns, M.; Gräter, F. *PLoS Comput. Biol.* **2009**, *5*, e1000306.
(31) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. *Biophys. J.* **2009**, *96*, 3997−4005.
(32) Stacklies, W.; Seifert, C.; Graeter, F. *BMC Bioinf.* **2011**, *12*, 101.
(33) Seifert, C.; Gräter, F. *Biophys. J.* **2012**, *103*, 2195−2202.
(34) Palmai, Z.; Seifert, C.; Gräter, F.; Balog, E. *PLoS Comput. Biol.* **2014**, *10*, e1003444.
(35) Salama-Alber, O.; Jobby, M. K.; Chitayat, S.; Smith, S. P.; White, B. A.; Shimon, L. J. W.; Lamed, R.; Frolow, F.; Bayer, E. A. *J. Biol. Chem.* **2013**, *288*, 16827−16838.
(36) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. *Biophys. J.* **2008**, *95*, 426−434.
(37) Friddle, R. W.; Noy, A.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13573−13578.
(38) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(39) Greene, R. F.; Callen, H. B. *Phys. Rev.* **1951**, *83*, 1231−1235.
(40) Erman, B. *Phys. Biol.* **2011**, *8*, 056003.
(41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(42) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(43) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(44) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127−1130.
(45) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.
(46) Jobst, M. A.; Schoeler, C.; Malinowska, K. H.; Nash, M. A. *J. Visualized Exp.* **2013**, *82*, e50950.
(47) Wojciechowski, M.; Thompson, D.; Cieplak, M. *J. Chem. Phys.* **2014**, *141*, 245103.
(48) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. *Biochemistry* **2007**, *46*, 2156−2173.
(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Appendix D

Article in Physical Review E
Chapter 5 in original form

Biasing effects of receptor-ligand complexes un protein-unfolding statistics

by

Constantin Schoeler, Tobias Verdorfer, Hermann E. Gaub, and Michael A. Nash

published in

Physical Review E 94, 042412 (2016)
Reprinted from Schoeler *et al.* [80] with permission from the American Physical Society

PHYSICAL REVIEW E **94**, 042412 (2016)

Biasing effects of receptor-ligand complexes on protein-unfolding statistics

Constantin Schoeler,[1] Tobias Verdorfer,[1] Hermann E. Gaub,[1] and Michael A. Nash[1,2,3,*]

[1]*Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 Munich, Germany*

[2]*Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland*

[3]*Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich (ETH-Zürich), 4058 Basel, Switzerland*

(Received 30 May 2016; published 13 October 2016)

Protein receptor-ligand pairs are increasingly used as specific molecular handles in single-molecule protein-unfolding experiments. Further, known marker domains, also referred to as fingerprints, provide unique unfolding signatures to identify specific single-molecule interactions, when receptor-ligand pairs themselves are investigated. We show here that in cases where there is an overlap between the probability distribution associated with fingerprint domain unfolding and that associated with receptor-ligand dissociation, the experimentally measured force distributions are mutually biased. This biasing effect masks the true parameters of the underlying free energy landscape. To address this, we present a model-free theoretical framework that corrects for the biasing effect caused by such overlapping distributions.

DOI: 10.1103/PhysRevE.94.042412

I. INTRODUCTION

Mechanical forces play an important role in many biological systems. Refolding of individual titin domains is believed to assist in muscle contraction [1], stretching forces expose cryptic binding sites involved in focal adhesions [2], and mechanically stable receptor-ligand pairs govern the assembly of large extracellular machineries and adhesion of bacterial cells to their cellulosic carbon sources [3,4]. Single-molecule pulling experiments with atomic force microscopes [5], optical tweezers [6], or magnetic tweezers [7] have become widely used techniques to study such phenomena at the single-molecule level.

Due to the stochastic nature of domain unfolding, typical atomic force microscopy experiments record many thousands of data traces to obtain statistically meaningful results from single-molecule pulling experiments. To unambiguously identify the unfolding signals of a given protein domain of interest or the dissociation of a receptor-ligand system under external load, the resulting data sets need to be filtered, and specific, single-molecule interactions must be identified.

To accommodate this need, the community has developed an elegant strategy to achieve high yields of specific curves: In this approach, protein domains of interest are fused to a receptor complex that serves as a specific handle in pulling experiments. This improves curve yields and data fidelity by providing a specific molecular interaction handle to "grab" the protein of interest. Thereby, the unfolding of individual domains and the dissociation of a receptor-ligand complex can be studied in a single experiment [3,4,8–10]. Also, using a known protein domain in the fusion construct provides a unique unfolding pattern that can be used to identify specific traces, when receptor-ligand unbinding itself is studied. These domains are then referred to as fingerprint domains [5].

In order for a curve to be unambiguously identified as constituting specific signal, it needs to exhibit unfolding of all included fingerprint domains plus a specific receptor-ligand dissociation event. Throughout this letter we will refer to domains fused to a receptor-ligand complex as fingerprint domains for both scenarios, namely protein-unfolding studies using receptor-ligand complexes as specific handles, as well as unbinding studies of receptor-ligand complexes of interest, which use fingerprint domains for assistance in data filtering. We discuss the statistical effects that arise when the respective force distributions for fingerprint domain unfolding and receptor-ligand complex rupture exhibit a finite overlap. We quantitatively show how the statistical overlap between these two distributions affects the experimentally observable unfolding and rupture force distributions. We provide a framework for extracting kinetic and energetic information from the experimentally observed distributions that are corrected for the biasing effects arising from such overlaps in a model-free fashion.

II. THEORETICAL FRAMEWORK

The standard theoretical framework treats protein unfolding or bond dissociation as thermally driven escape over a free energy barrier that is modulated by an external force F [11–14]. This description leads to a general expression for the distribution of unfolding or rupture forces in a pulling experiment,

$$p(F) = \frac{k(F)}{\dot{F}(F)} \exp\left(-\int_0^F \mathrm{d}F' \, \frac{k(F')}{\dot{F}(F')}\right), \tag{1}$$

where $k(F)$ is the force-dependent off rate of the system, and $\dot{F}(F)$ is the force loading rate. In the simplest picture [12,13], the force-dependent off rate is given by [11]

$$k(F) = k_0 \exp\left(\frac{F\Delta x}{k_\mathrm{B}T}\right) \tag{2}$$

where k_0 and Δx are the zero-force off rate and distance to the free energy barrier, respectively. For a constant force loading rate $\dot{F}$ and an off rate from Eq. (2), the integral in Eq. (1) can be solved analytically [Fig. 1(a)]. Dudko

*michael.nash@unibas.ch

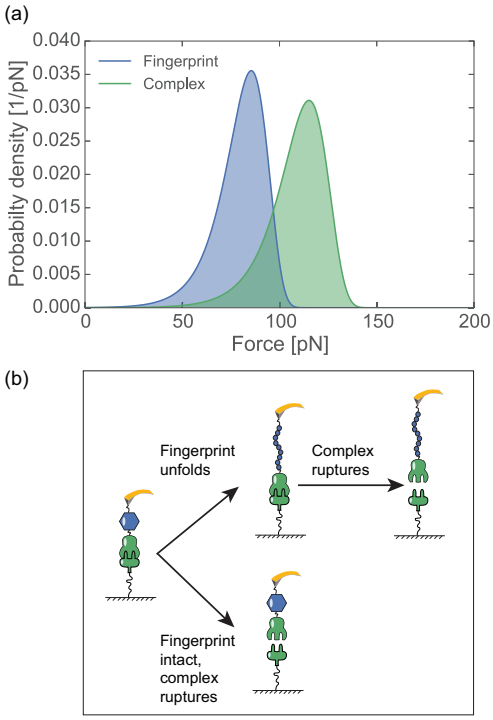


FIG. 1. (a) Overlapping distributions of fingerprint unfolding (blue region) and complex rupture (green region) for a constant loading rate $\dot{F} = 200$ pN s^{-1} with fingerprint $\Delta x = 0.4$ nm, and $k_0 = 0.005$ s^{-1} and complex parameters $\Delta x = 0.35$ nm, and $k_0 = 0.001$ s^{-1} obtained by integrating Eq. (1) using Eq. (2). (b) Schematic of possible outcomes for the situation in (a). Analyzable data show fingerprint unfolding followed by complex rupture (upper path). Due to the overlapping distributions for unfolding and rupture, complex rupture with an intact fingerprint is also possible (lower path).

et al. [15] have used the Kramers theory [16] to obtain a more sophisticated expression for the force-dependent off rate, which accounts for the specific shape of the free energy landscape. This approach also provides an analytical solution to Eq. (1) for a constant loading rate and includes the height of the free energy barrier ΔG as an additional parameter. Over the years, more theoretical models describing the distributions for domain unfolding and receptor-ligand dissociation have been developed [17–22] that can be applied to experimentally measured force distributions.

Since a constant force loading rate is required to obtain an analytical solution for the distribution of forces in a pulling experiment, force ramp measurements [23,24], where the external force is ramped linearly, are an elegant technique to study protein unfolding and receptor-ligand dissociation. In an experiment where a receptor-ligand system is used to probe the unfolding behavior of a protein fingerprint domain of interest, care has to be taken when analyzing the resulting unfolding or rupture force distributions. If the probability distributions for fingerprint domain unfolding and complex rupture are disjunct, i.e., the complex ruptures at forces well above those required for fingerprint unfolding, the measured distributions are unbiased and can be readily analyzed using a desired form of Eq. (1). If those distributions have a substantial overlap [Fig. 1(a)], however, biasing effects have to be taken into account.

To pass the data analysis filter a given curve is required to exhibit both fingerprint unfolding and complex rupture [Fig. 1(b); upper path], i.e., the complex must not rupture prior to fingerprint unfolding [Fig. 1(b); lower path]. Consequently, the resulting distribution of fingerprint unfolding forces will be shifted downwards towards lower forces. By the same logic, experimentally observed distributions for receptor-ligand complex rupture forces will be truncated at the lower end and shifted upwards toward higher forces in a constant-loading-rate experiment. This biasing effect has been used qualitatively by Jobst et al. [8] to unambiguously identify a redundant dual binding mode in a receptor-ligand complex.

Here we treat this biasing effect quantitatively and calculate these effects independent of the model used in Eq. (1). For overlapping distributions of fingerprint, $p_\mathrm{f}(F)$, and receptor-ligand complex, $p_\mathrm{c}(F)$, the biased distribution of the fingerprint, $p_\mathrm{f}^\star(F)$, can be calculated by multiplying the original distribution by the cumulative probability for the bond to rupture at higher forces and renormalizing,

$$p_\mathrm{f}^\star(F) = \frac{p_\mathrm{f}(F)\int_F^\infty \mathrm{d}F'\, p_\mathrm{c}(F')}{\eta}, \tag{3}$$

where η is a normalization constant. Since $p_\mathrm{c}(f)$ is normalized, Eq. (3) can be rewritten,

$$p_\mathrm{f}^\star(F) = \frac{p_\mathrm{f}(F)\left(1 - \int_0^F \mathrm{d}F'\, p_\mathrm{c}(F')\right)}{\eta} \tag{4}$$

$$= \frac{p_\mathrm{f}(F)(1 - P_\mathrm{c}(F))}{\eta}, \tag{5}$$

where $P_\mathrm{c}(F)$ is the cumulative distribution function of the complex rupture forces. The normalization constant η can be calculated by integrating over the numerator in Eq. (5):

$$\eta = 1 - \int_0^\infty \mathrm{d}F\, p_\mathrm{f}(F) P_\mathrm{c}(F). \tag{6}$$

Intuitively, the biased fingerprint distribution is normalized by the ratio of curves that exhibit fingerprint unfolding vs all rupture events. The above calculations apply for both constant-loading-rate and constant-speed measurements. By the same logic, the biased distribution of observed complex ruptures for a constant loading rate can be calculated as

$$p_\mathrm{c}^\star(F) = \frac{p_\mathrm{c}(F) P_\mathrm{f}(F)}{\eta}. \tag{7}$$

Both biased distributions for fingerprint unfolding, $p_\mathrm{f}^\star(F)$, and complex ruptures, $p_\mathrm{c}^\star(F)$, are normalized by the same yield parameter η since both distributions are extracted from the same curves in a given data set. For a mathematical proof, see Appendix A. We note that the biasing effect on complex rupture in the constant-speed case is more difficult to quantify.

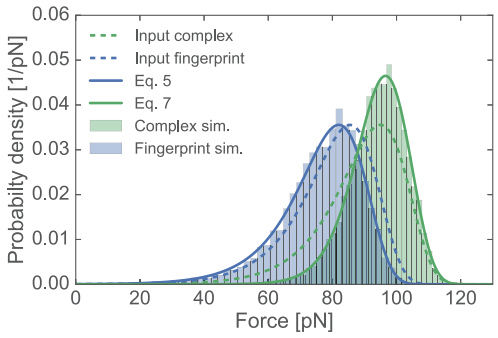


FIG. 2. Force ramp simulation with $\dot{F} = 200$ pN s^{-1}, complex $\Delta x_c = 0.4$ nm, $k_{0,c} = 0.002$ s^{-1}, and fingerprint $\Delta x_f = 0.4$ nm, $k_{0,f} = 0.005$ s^{-1} ($\eta = 0.71$). Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

Since the additional contour length released upon fingerprint unfolding is not immediately compensated for by a feedback on the force signal, there is a pronounced drop in force associated with fingerprint unfolding, giving rise to the characteristic sawtooth pattern in force extension traces. Ignoring the finite relaxation time of the harmonic pulling device, the force will drop from $F_1 = F(x,L)$ to $F_2 = F(x,L+\Delta L)$, where the former describes the (nonlinear) elastic behavior of the pulling device and potential linker molecules, and ΔL is the additional contour length released upon fingerprint unfolding. Whether or not the observed complex distribution is biased depends on the value of $p_c(F_2)$. For $p_c(F_2) \approx 0$, no biasing will occur, whereas $p_c(F_2) > 0$ will cause a biasing effect. In practice, ΔL is usually large enough to ensure that complex distributions are unbiased in constant-speed experiments, leaving a substantial bias only on the observed fingerprint distribution. A strategy to implement our correction procedure for the constant-speed protocol is proposed in Appendix B.

III. MONTE CARLO SIMULATIONS

To validate our results, we used a Monte Carlo approach to simulate fingerprint domain unfolding in combination with receptor-ligand dissociation in a constant-loading-rate protocol. Our simulation routine is similar to the approach described in Ref. [25] and uses the phenomenological model due to its analytical tractability. Briefly, we integrate Eq. (2) over a time step Δt, where $F = F(t) = \dot{F}t$, to obtain probabilities p_u and p_r for fingerprint unfolding and complex rupture, respectively. These probabilities are compared to independent random numbers n_u and n_r between 0 and unity. If $p_u < n_u$ and $p_r < n_r$, the fingerprint and complex remain intact and the next iteration is started. If $p_u > n_u$ and $p_r < n_r$, the corresponding force is noted as the fingerprint unfolding force and the next iteration is started with only the complex intact. The simulation then continues until $p_r > n_r$ and the corresponding force is noted as the complex rupture force. If $p_u < n_u$ and $p_r > n_r$ or $p_u > n_u$ and $p_r > n_r$, the complex ruptured prior to fingerprint unfolding or at the same time, and an experimental curve would be unanalyzable and filtered out during the data analysis procedure.

Results from a Monte Carlo simulation at constant loading rate $\dot{F} = 200$ pN s^{-1} with overlapping distributions for fingerprint unfolding and complex rupture are shown in Fig. 2. $N = 10000$ traces were simulated and the observed fingerprint domain unfolding forces and receptor-ligand complex rupture forces were histogrammed. We only analyzed curves that showed both fingerprint unfolding and complex rupture to mimic experimental conditions. As expected, both fingerprint unfolding and complex rupture distributions (blue and green histograms in Fig. 2) are shifted from the unbiased input distributions (dashed blue and green lines in Fig. 2). The biased results are well described by our theoretical predictions, which are shown as solid blue and green lines in Fig. 2. To illustrate the potential errors that can occur from not accounting for the fingerprint biasing effect, we used the uncorrected distribution [Eq. (1)] to fit the biased fingerprint domain unfolding and complex rupture histograms and compared the resulting fit parameters Δx_{fit} and $k_{0,fit}$ with the unbiased input parameters (Table I). We found that for fingerprint domain unfolding, Δx is hardly affected, while k_0 is overestimated by 30%. For complex rupture Δx is overestimated by 29% due to the smaller width of the biased distribution, while k_0 is underestimated by over an order of magnitude. If the unbiased parameters for the complex distribution are known from a control experiment, our predicted biased distributions can be used to fit the experimental data to obtain unbiased values for the fit parameters pertaining to the fingerprint, or vice versa. Alternatively, a global fitting procedure can be applied to both biased distributions for constant-loading-rate data to obtain unbiased fit parameters without prior knowledge of either distribution. Using this approach, we obtained global fit parameters that were within 6% of the input parameters (Table I).

TABLE I. Input vs fit parameters of the simulation shown in Fig. 2. For these parameters, the yield parameter equals $\eta = 0.71$. Initially, the simulated distributions were fit with the uncorrected distributions ($k_{0,fit}$ and Δx_{fit}). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions [Eqs. (5) and (7)] in a global fitting procedure to obtain the corrected parameters $k_{0,global}$ and Δx_{global}.

	$k_{0,input}$ (s^{-1})	$k_{0,fit}$ (s^{-1})	$k_{0,global}$ (s^{-1})	Δx_{input} (nm)	Δx_{fit} (nm)	Δx_{global} (nm)
Fingerprint	5.0×10^{-3}	$(6.6 \pm 0.2) \times 10^{-3}$	$(4.7 \times 0.4) \times 10^{-3}$	0.400	0.402 ± 0.007	0.401 ± 0.005
Complex	2.0×10^{-3}	$(0.10 \pm 0.02) \times 10^{-3}$	$(1.9 \pm 0.02) \times 10^{-3}$	0.400	0.527 ± 0.007	0.402 ± 0.005

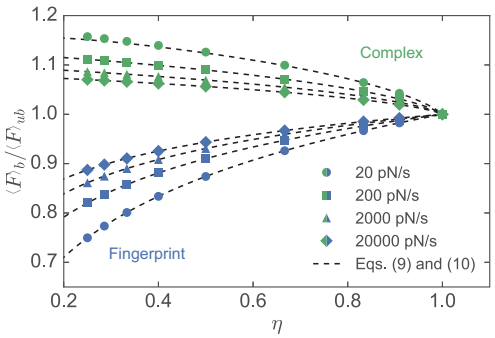


FIG. 3. Biasing of fingerprint unfolding and complex rupture forces as a function of the theoretical yield of curves exhibiting fingerprint unfolding η. Data points show the relative mean biased unfolding force $\langle F\rangle_b/\langle F\rangle_{ub}$; fingerprint and complex data are shown in blue and green, respectively. For fingerprint data, parameters were held constant at $\Delta x_f = 0.4$ nm and $k_{0,f} = 0.005\ s^{-1}$, while the complex distribution was shifted by maintaining $\Delta x_c = 0.4$ nm and varying the off rate. For complex data, $\Delta x_c = 0.4$ nm and $k_{0,c} = 0.005\ s^{-1}$ were held constant and the fingerprint distribution was shifted by maintaining $\Delta x_f = 0.4$ nm and varying the off rate. Dashed lines represent predictions based on Eq. (9) and (10).

IV. SPECIAL CASE: EQUAL POTENTIAL WIDTHS

In Fig. 3 we quantify the magnitude of the biasing effect by numerically calculating the relative mean biased unfolding force of the fingerprint (blue symbols) and complex (green symbols) $\langle F\rangle_b/\langle F\rangle_{ub} = \int dF\, F p^*_{f/c}(F)/\int dF\, F p_{f/c}(F)$ as a function of the theoretical ratio of curves exhibiting fingerprint unfolding prior to receptor-ligand complex rupture, η [Eq. (6)]. For analytical tractability we chose the special case $\Delta x_c = \Delta x_f$. With this simplification we find for the fingerprint

$$\langle F\rangle_b = \frac{k_B T}{\Delta x_f} e^{\frac{k_B T k_{0,f}}{\Delta x_f r \eta}} E_1\left(\frac{k_B T k_{0,f}}{\Delta x_f r \eta}\right), \tag{8}$$

where $E_1(x)$ is the exponential integral. Using $e^x E_1(x) \cong \ln(1+\frac{e^{-\gamma}}{x})$ we can produce an analytical approximation for the relative mean biased unfolding force for the aforementioned special case,

$$\frac{\langle F\rangle_b}{\langle F\rangle_{ub}} \cong \frac{\ln\left(1+\frac{r\Delta x_f}{k_{0,f}k_B T}e^{-\gamma}\eta\right)}{\ln\left(1+\frac{r\Delta x_f}{k_{0,f}k_B T}e^{-\gamma}\right)}, \tag{9}$$

where $\gamma = 0.577\ldots$ is the Euler-Mascheroni constant. The analogous result for the complex reads

$$\frac{\langle F\rangle_b}{\langle F\rangle_{ub}} \cong \frac{1}{\eta} - \frac{\ln\left(1+\frac{r\Delta x_c}{k_{0,c}k_B T}e^{-\gamma}(1-\eta)\right)}{\ln\left(1+\frac{r\Delta x_c}{k_{0,c}k_B T}e^{-\gamma}\right)}\frac{1-\eta}{\eta}. \tag{10}$$

For this special case Eq. (6) can be evaluated analytically, yielding $\eta = (1+\frac{k_{0,c}}{k_{0,f}})^{-1}$. Equations (9) and (10) (dashed lines in Fig. 3) agree very well with our numerical results over the loading-rate regime covered. Figure 3 clearly shows that the biasing effect is more pronounced for low loading rates, consistent with our theoretical predictions based on Eqs. (9) and (10). In cases where data cannot be corrected for a potential biasing effect, e.g., due to low experimental yields, the magnitude of the biasing effect can be minimized by increasing the loading rate in a pulling experiment.

V. CONCLUSION

Our calculations provide a framework for analyzing single-molecule force spectroscopy data where receptor-ligand systems are used as specific handles to study a fingerprint domain of interest, or vice versa. In such experiments, it is many times the case that the distributions of fingerprint domain unfolding and complex rupture have a significant overlap (a few exemplary cases can be found in Refs. [3,9,10,26]). In this case biasing effects will occur and should be considered in the analysis procedure. Our findings can be applied to both constant-speed and force ramp (constant-loading-rate) experimental protocols, however, it should be noted that the biasing effect on complex unbinding is more complicated in the constant-speed protocol, due to the drop in force upon fingerprint unfolding dependent on the length of the unfolded domain. Since the biasing effects are solely due to the statistical nature of domain unfolding or complex unbinding, our results, specifically Eqs. (5) and (7), are valid regardless of the specific model used in Eq. (1).

ACKNOWLEDGMENTS

The authors would like to thank Ellis Durner, Markus A. Jobst, Magnus S. Bauer, Fabian Baumann, and Steffen Sedlak for valuable discussions. This work was supported by an advanced grant to H.E.G. from the European Research Council (Cellufuel Grant No. 294438). M.A.N. acknowledges funding from Society in Science The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland.

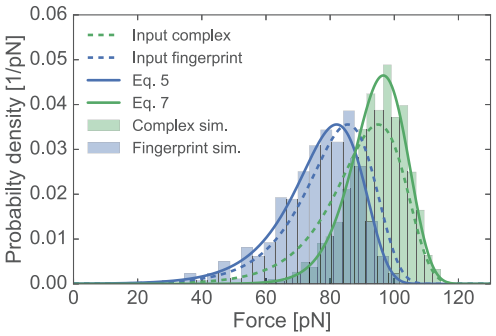


FIG. 4. Force ramp simulation with $\dot{F} = 200\ pN\,s^{-1}$, complex $\Delta x_c = 0.4$ nm, $k_{0,c} = 0.002\ s^{-1}$, fingerprint $\Delta x_f = 0.4$ nm, $k_{0,f} = 0.005\ s^{-1}$ ($\eta = 0.71$), and simulated traces $N = 1000$. Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

TABLE II. Input vs fit parameters of the simulation shown in Fig. 4. For these parameters, the yield parameter equals $\eta = 0.71$. Initially, the simulated distributions were fit with the uncorrected distributions ($k_{0,\mathrm{fit}}$ and Δx_{fit}). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions [Eqs. (5) and (7)] in a global fitting procedure to obtain the corrected parameters $k_{0,\mathrm{global}}$ and $\Delta x_{\mathrm{global}}$.

	$k_{0,\mathrm{input}}$ (s^{-1})	$k_{0,\mathrm{fit}}$ (s^{-1})	$k_{0,\mathrm{global}}$ (s^{-1})	$\Delta x_{\mathrm{input}}$(nm)	Δx_{fit} (nm)	$\Delta x_{\mathrm{global}}$ (nm)
Fingerprint	5.0×10^{-3}	$(5.9 \pm 1.2) \times 10^{-3}$	$(4.3 \pm 1.5) \times 10^{-3}$	0.400	0.409 ± 0.011	0.410 ± 0.020
Complex	2.0×10^{-3}	$(0.2 \pm 0.1) \times 10^{-3}$	$(2.7 \pm 1.3) \times 10^{-3}$	0.400	0.504 ± 0.020	0.384 ± 0.021

APPENDIX A: IDENTITY OF YIELD PARAMETER η IN EQS. (5) AND (7)

As we state in the text, the normalization parameter η is equal to the ratio of curves that exhibit fingerprint unfolding vs all rupture events. In other words a fraction $1-\eta$ of all curves will be "missed" in an actual experiment, since they do not exhibit fingerprint unfolding and would hence be discarded during data analysis. Consequently, both fingerprint and complex distributions (which are obtained from the same curves) need to be normalized by the same η to obtain probability distributions normalized to unity. Mathematically, one needs to show that

$$1 - \int_0^\infty \mathrm{d}F\; p_{\mathrm{f}}(F) P_{\mathrm{c}}(F) = \eta = \int_0^\infty \mathrm{d}F\; p_{\mathrm{c}}(F) P_{\mathrm{f}}(F). \quad \text{(A1)}$$

We use integration by parts and $\int_0^F \mathrm{d}F'\, p(F') = P(F)$ to evaluate the right-hand side (rhs) of Eq. (A1):

$$\mathrm{rhs} = P_{\mathrm{c}}(F) P_{\mathrm{f}}(F)\big|_0^\infty - \int_0^\infty \mathrm{d}F\; P_{\mathrm{c}}(F) p_{\mathrm{f}}(F). \quad \text{(A2)}$$

Since $P(0) = 0$ and $P(\infty) = 1$ this is equal to the left-hand side of Eq. (A1).

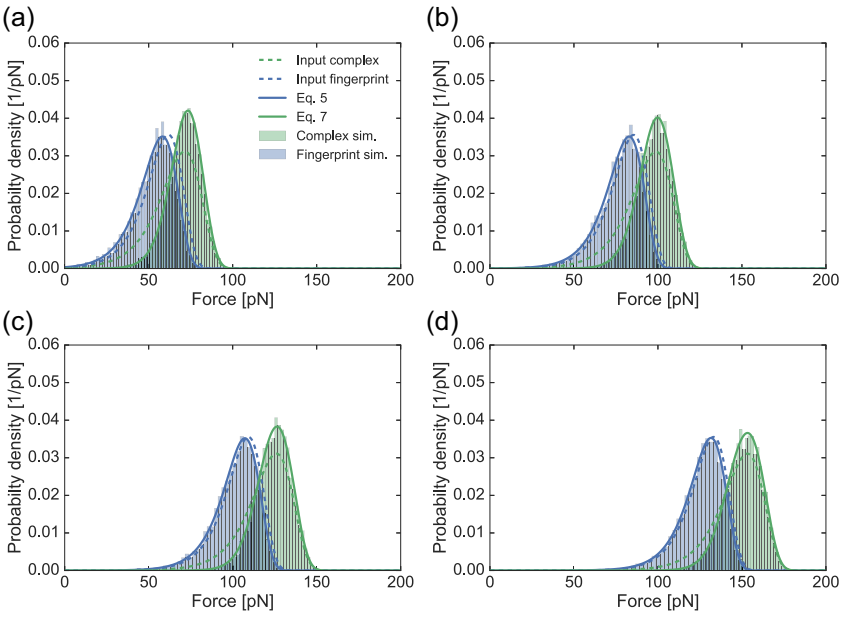


FIG. 5. Force ramp simulation with complex $\Delta x_{\mathrm{c}} = 0.35$ nm, $k_{0,\mathrm{c}} = 0.004\ \mathrm{s}^{-1}$, fingerprint $\Delta x_{\mathrm{f}} = 0.4$ nm, $k_{0,\mathrm{f}} = 0.005\ \mathrm{s}^{-1}$, and varying $\dot{F}$. Due to the different potential widths Δx, the yield parameter η changes with the loading rate. Values of the loading rate and resulting yield parameter are (a) $\dot{F} = 20\ \mathrm{pN\,s}^{-1}$ and $\eta = 0.70$, (b) $\dot{F} = 200\ \mathrm{pN\,s}^{-1}$ and $\eta = 0.75$, (c) $\dot{F} = 2000\ \mathrm{pN\,s}^{-1}$ and $\eta = 0.80$, and (d) $\dot{F} = 20000\ \mathrm{pN\,s}^{-1}$ and $\eta = 0.85$. Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

APPENDIX B: CONSTANT SPEED

As pointed out in the text, the correction procedure for distributions obtained from constant-speed measurements is more involved. In this Appendix we discuss a strategy for extracting unbiased parameters from fingerprint distributions for this pulling protocol. Due to the nonlinear elasticity of linker molecules (e.g., polyethylene glycol, spacers, or unfolded protein backbone), the force loading rate $\dot{F} = \dot{F}(F)$ becomes a function of the force and the integral in Eq. (1) can no longer be evaluated analytically. In a standard pulling experiment, a harmonic pulling device (e.g., atomic force microscopy cantilever or optically trapped bead) is connected to the aforementioned linker molecules. By applying a force balance it can be shown that the force loading rate is given by [14]

$$\frac{v}{\dot{F}(F)} = \frac{1}{k_\mathrm{h}} + \frac{\partial x(F)}{\partial F}, \tag{B1}$$

where v is the pulling speed, k_h is the spring constant of the harmonic pulling device, and $x(F)$ is the force-dependent extension of the linker. The biased distribution for fingerprint unfolding $p_\mathrm{f}^*(F)$ can be computed by numerically solving the integrals occurring in Eqs. (5) and (6), using a model for the elastic response of the linker molecules such as the worm-like-chain model, freely rotating model, or composite model proposed by Livadaru et al. [27] in Eq. (B1). The choice of model should be made based on the force regime in which fingerprint unfolding and complex rupture are expected and the molecular linkers present in an experimental setup need to be accounted for in these models via their contour length, L, and elasticity, e.g., persistence length p. The force-dependent loading rate for the worm-like-chain model has been derived as Eq. (4) in Ref. [14].

APPENDIX C: SUPPLEMENTAL FIGURES AND TABLES

This Appendix contains contains two figures and one table that support the results in the text. Figure 4 shows the results of a simulation with parameters identical to those in Fig. 2, except only $N = 1000$ traces were simulated, to mimic a total number of curves more similar to average experimental yields. The extracted fit parameters for this simulation are listed in Table II. We note that despite the increase in uncertainties, the extracted parameters from our global fitting procedure still reproduce the input parameters much better than those obtained from a fit to the uncorrected distributions. Figure 5 shows the results of Monte Carlo simulations at different loading rates ranging from $\dot{F} = 20\ \mathrm{pN\,s^{-1}}$ to $\dot{F} = 20\,000\ \mathrm{pN\,s^{-1}}$. Due to the different potential widths $\Delta x_\mathrm{c} = 0.35$ nm and $\Delta x_\mathrm{f} = 0.4$ nm, the yield parameter η varies for the different simulations.

[1] J. A. Rivas-Pardo, E. C. Eckels, I. Popa, P. Kosuri, W. A. Linke, and J. M. Fernández, Cell Reports **14**, 1339 (2016).
[2] A. del Rio, R. Perez-Jimenez, R. Liu, P. Roca-Cusachs, J. M. Fernandez, and M. P. Sheetz, Science **323**, 638 (2009).
[3] S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, and H. E. Gaub, Proc. Natl. Acad. Sci. USA **109**, 20431 (2012).
[4] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, Nature Commun. **5**, 5635 (2014).
[5] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, J. Struct. Biol. (2016), doi: 10.1016/j.jsb.2016.02.011.
[6] K. C. Neuman and A. Nagy, Nature Methods **5**, 491 (2008).
[7] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[8] M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, eLife **4**, e10319 (2015).
[9] F. Baumann, M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, and D. A. Pippig, Nature Nanotechnol. **11**, 89 (2016).
[10] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub, Nature Methods **11**, 1127 (2014).
[11] G. Bell, Science **200**, 618 (1978).
[12] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, Biophys. J. **72**, 1568 (1997).
[13] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[14] O. K. Dudko, G. Hummer, and A. Szabo, Proc. Natl. Acad. Sci. USA **105**, 15755 (2008).
[15] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[16] H. A. Kramers, Physica **7**, 284 (1940).
[17] O. K. Dudko, Q. Rev. Biophys. **49**, e3 (2015).
[18] B. Heymann and H. Grubmüller, Phys. Rev. Lett. **84**, 6126 (2000).
[19] J. T. Bullerjahn, S. Sturm, and K. Kroy, Nature Commun. **5**, 4463 (2014).
[20] R. W. Friddle, Phys. Rev. Lett. **100**, 138302 (2008).
[21] R. W. Friddle, A. Noy, and J. J. De Yoreo, Proc. Natl. Acad. Sci. USA **109**, 13573 (2012).
[22] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, Phys. Rev. Lett. **94**, 048301 (2005).
[23] A. F. Oberhauser, P. K. Hansma, M. Carrión-Vázquez, and J. M. Fernandez, Proc. Natl. Acad. Sci. USA **98**, 468 (2001).
[24] E. Evans, A. Leung, V. Heinrich, and C. Zhu, Proc. Natl. Acad. Sci. USA **101**, 11281 (2004).
[25] C. Friedsam, A. K. Wehle, F. Kühner, and H. E. Gaub, J. Phys.: Condens. Matter **15**, S1709 (2003).
[26] J. P. Mueller, S. Mielke, A. Loef, T. Obser, C. Beer, L. K. Bruetzel, D. A. Pippig, W. Vanderlinden, J. Lipfert, R. Schneppenheim, and M. Benoit, Proc. Natl. Acad. Sci. USA **113**, 1208 (2016).
[27] L. Livadaru, R. R. Netz, and H. J. Kreuzer, Macromolecules **36**, 3732 (2003).

Appendix E

Research Article in eLife

Resolving Dual Binding Conformations of Cellulosome Cohesin-Dockerin Complexes Using Single-Molecule Force Spectroscopy

by

Markus A. Jobst, Lukas F. Milles, Constantin Schoeler, Wolfgang Ott, Daniel B. Fried, Edward A. Bayer, Hermann E. Gaub, and Michael A. Nash

published in

eLife 4:e1031. doi: 10.7554/eLife.10319 (2015)

eLIFE elifesciences.org

RESEARCH ARTICLE

Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy

Markus A Jobst[1,2], Lukas F Milles[1,2], Constantin Schoeler[1,2], Wolfgang Ott[1,2], Daniel B Fried[3], Edward A Bayer[4], Hermann E Gaub[1,2], Michael A Nash[1,2]*

[1]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-University, Munich, Germany; [2]Center for Nanoscience, Ludwig-Maximilians-University, Munich, Germany; [3]Kean University, New Jersey, United States; [4]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel

Abstract Receptor-ligand pairs are ordinarily thought to interact through a lock and key mechanism, where a unique molecular conformation is formed upon binding. Contrary to this paradigm, cellulosomal cohesin-dockerin (Coh-Doc) pairs are believed to interact through redundant dual binding modes consisting of two distinct conformations. Here, we combined site-directed mutagenesis and single-molecule force spectroscopy (SMFS) to study the unbinding of Coh:Doc complexes under force. We designed Doc mutations to knock out each binding mode, and compared their single-molecule unfolding patterns as they were dissociated from Coh using an atomic force microscope (AFM) cantilever. Although average bulk measurements were unable to resolve the differences in Doc binding modes due to the similarity of the interactions, with a single-molecule method we were able to discriminate the two modes based on distinct differences in their mechanical properties. We conclude that under native conditions wild-type Doc from *Clostridium thermocellum* exocellulase Cel48S populates both binding modes with similar probabilities. Given the vast number of Doc domains with predicteddual binding modes across multiple bacterial species, our approach opens up newpossibilities for understanding assembly and catalytic properties of a broadrange of multi-enzyme complexes.
DOI: 10.7554/eLife.10319.001

***For correspondence:** michael.nash@lmu.de

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 23 July 2015
Accepted: 28 October 2015
Published: 31 October 2015

Reviewing editor: Taekjip Ha, Johns Hopkins University School of Medicine, United States

Introduction

Cellulosomes are hierarchically branching protein networks developed by nature for efficient deconstruction of lignocellulosic biomass. These enzyme complexes incorporate catalytic domains, carbohydrate binding modules (CBMs), cohesin:dockerin (Coh:Doc) pairs, and other conserved features (***Demain et al., 2005***; ***Bayer et al., 2004***; ***Schwarz, 2001***; ***Béguin and Aubert, 1994***; ***Smith and Bayer, 2013***; ***Fontes and Gilbert, 2010***). A central attribute of cellulosome assembly is the conserved ~75 amino acid type-I Doc domain typically found at the C-terminus of cellulosomal catalytic domains. The highly conserved consensus Doc sequence from *Clostridium thermocellum (Ct)* is shown in ***Figure 1A***. Dockerins guide attachment of enzymes into the networks by binding strongly to conserved Coh domains organized within non-catalytic poly (Coh) scaffolds. In addition to their nanomolar binding affinities, many archetypal Coh:Doc pairs are thought to exhibit dual binding modes (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***; ***Currie et al., 2012***). The bound Doc domain can adopt two possible orientations that differ by ~180° rotation on the Coh surface, as shown in ***Figure 1B***. The two binding modes originate from duplicated F-hand sequence motifs, a conserved structural feature found among type-I dockerins (***Pagès et al., 1997***). The duplicated F-hand motifs resemble EF-hands found in eukaryotic calcium binding proteins (e.g., calmodulin), and provide

eLife digest Some bacteria use cellulose, the main component of plant cell walls, as a food source. The enzymes that break down cellulose are anchored onto a protein scaffold in a structure called the cellulosome on the bacteria's surface. This anchoring occurs through an interaction between receptor proteins known as 'cohesin' domains on the scaffold proteins and 'dockerin' ligands on the enzymes.

Most receptor-ligand interactions only allow the two proteins to bind in a single, fixed orientation. However, cohesins and dockerins are suspected to bind in two different configurations. It has been difficult to investigate the populations of these different configurations because most experimental techniques investigating protein binding take average measurements from many molecules at once. As the binding modes are extremely similar, these methods have been unable to distinguish between the two cohesin-dockerin binding configurations without introducing mutations, in part because these configurations are very similar to each other.

Jobst et al. used a technique called single-molecule force spectroscopy to investigate cohesin-dockerin interactions between individual molecules. This technique applies a force that separates, or 'unbinds', cohesin and dockerin, by pulling individual complexes of the two binding partners apart with a nanoscale probe. In the experiments, *E. coli* bacteria were made to produce mutant versions of dockerin that can only bind to cohesin in one orientation. This allowed each binding configuration to be studied individually. The results of these experiments revealed the mechanical unbinding patterns of each cohesin-dockerin configuration, and showed that it is possible to use these patterns to distinguish between the two configurations. A complimentary set of experiments revealed that wild-type (non-mutated) cohesin-dockerin complexes occupy both configurations in approximately equal amounts, and do not switch modes once bound.

Further single-molecule experiments together with computer simulations will provide a more detailed picture of how cohesin and dockerin fit together in the two configurations. Such experiments could also reveal how cohesin and dockerin contribute to the break down of cellulose inside living cells and how they could be used for the precise assembly of single proteins.

DOI: 10.7554/eLife.10319.002

internal sequence and structural symmetry to Doc domains. Rotating Doc by ~180° with respect to Coh (***Figure 1B,C***) results in an alternatively bound complex with similarly high affinity involving the same residues on Coh recognizing mirrored residues within Doc. The dual binding mode is thought to increase the conformational space available to densely packed enzymes on protein scaffolds, and to facilitate substrate recognition by catalytic domains within cellulosomal networks (***Bayer et al., 2004***). From an evolutionary perspective, the dual binding mode confers robustness against loss-of-function mutations, while allowing mutations within Doc to explore inter-bacterial species cohesin-binding promiscuity in cellulosome-producing microbial communities. Coh:Doc interactions and dual binding modes are therefore important in the context of cellulose degradation by cellulosome-producing anaerobic bacterial communities.

However, direct experimental observation of the dual binding modes for wild-type Doc has thus far proven challenging. Ensemble average bulk biochemical assays (e.g., surface plasmon resonance, calorimetry, enzyme-linked immunosorbent assays) are of limited use in resolving binding mode populations, particularly when the binding modes are of equal thermodynamic affinity. Crystallography is challenging because the complex does not adopt a unique molecular conformation, but rather exhibits a mixture of two conformations thereby hindering crystal growth. Structural data on the dual binding mode have typically been collected using a mutagenesis approach, where one of the binding modes was destabilized by mutating key recognition elements (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***). This approach, however, while resolving the structures of each bound complex, cannot determine if one binding mode is dominant for wild-type Doc, or if that dominance is species or sequence dependent. Coarse grained molecular dynamics has also predicted dual modes of interaction between Coh and Doc (***Hall and Sansom, 2009***), but direct experimental evidence of both binding modes for wild-type Doc has remained elusive. Improved fundamental understanding of the dual binding mode could shed light onto the molecular mechanisms by which these multi-

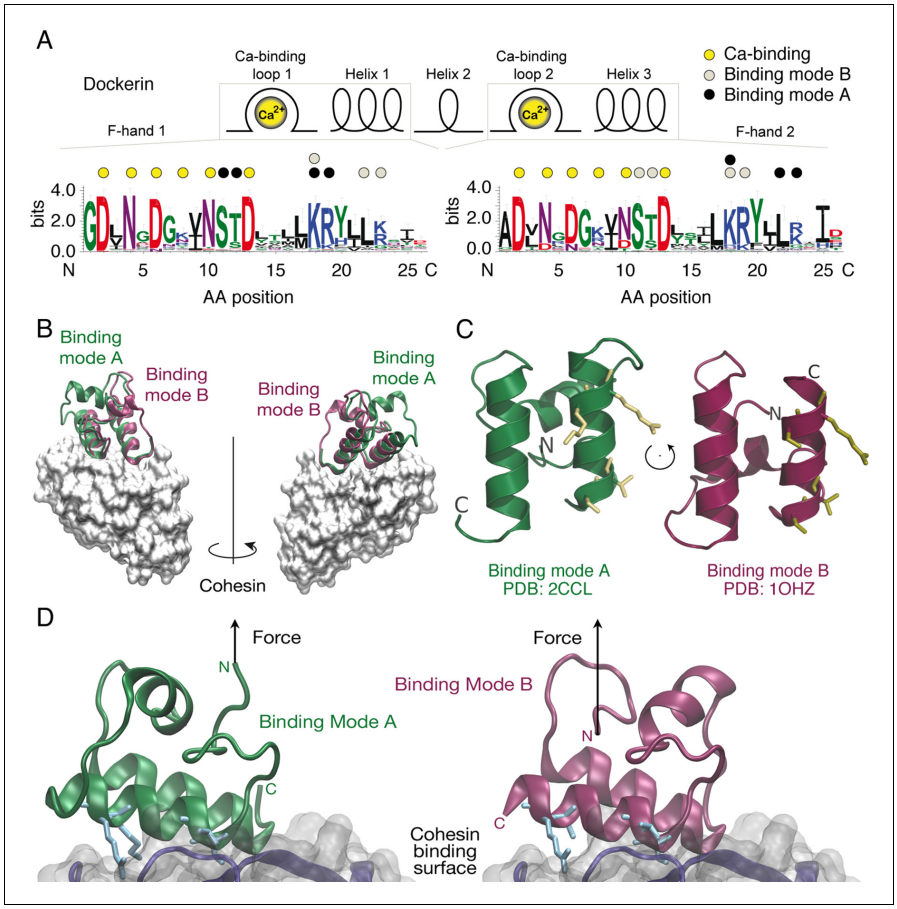


Figure 1. Cohesin:Dockerin dual binding modes. (**A**) Secondary structure and consensus sequence logo (***Crooks, 2004***) assembled from 65 putative *Ct* type-I Doc variants. Dots above the amino acid codes indicate residues involved in: Ca^{2+} coordination (yellow), mode A binding (black), and mode B binding (gray). Letter colors represent chemical properties: Green, polar; purple, neutral; blue, basic; red, acidic; black, hydrophobic. Crucial Coh-binding residues are located at positions 11, 12, 18, 19, 22, and 23 in each F-hand motif. (**B**) Coh:Doc complex crystal structures showing overlaid Doc domains in the two binding modes. Images were generated by aligning the Coh domain (gray) from PDB 2CCL (green, binding mode (**A**) and 1OHZ (red, binding mode (**B**) using the VMD plugin MultiSeq (***Humphrey et al., 1996***; ***Roberts et al., 2006***). (**C**) View of the Doc binding interface for each mode from the perspective of Coh. The conserved binding residues at positions 11, 12, 18, and 19 in the F-hand motif relevant for binding in the corresponding mode are depicted as stick models (yellow). (**D**) Close-up view of the interface for each binding mode with arrows indicating the location and direction of applied force. Binding residues 11, 12, 18, and 19 for binding mode A and 45, 46, 52, and 53 for binding mode B are shown as blue stick models. The Coh domain is oriented the exact same way in both views.

DOI: 10.7554/eLife.10319.003

enzyme complexes self-assemble and achieve synergistic conformations, as well as provide a new approach to designing systems for protein nanoassembly (***Kufer et al., 2009***; ***2008***).

Here, we used SMFS (***Li and Cao, 2010***; ***Engel and Müller, 2000***; ***Woodside and Block, 2014***) to study wild-type and mutant Doc from exocellulase Cel48S of *C. thermocellum* (*Ct*-DocS). We demonstrate that specific unfolding/unbinding trajectories of individually bound Coh:Doc complexes

are characteristic of the binding modes. To validate our approach, we produced Doc mutants that exhibited a preferred binding mode. We performed single-molecule pulling experiments on bound Coh:mutant Doc complexes and observed a strong bias in the probability of two clearly distinguishable unfolding patterns, termed 'single' and 'double' rupture types for each binding mode mutant. We further probed the unbinding mechanism of the double rupture events using poly (Gly-Ser) inserts to add amino acid sequence length to specific sections of Doc as a means to identify which portions of Doc unfolded. Finally, we used the inherent differences in mechanical stability of each binding mode, and the effects these differences had on the unfolding force distributions of an adjacent domain, to directly observe and quantify binding mode populations for wild-type Doc.

Results

Protein design

The wild-type and mutant Doc sequences used in this work were aligned (***Beitz, 2000***) and are presented in ***Figure 2***. Among *Ct*-Doc domains, a Ser-Thr pair located at positions 11 and 12 of F-hand motif 1 (N-terminal helix 1) is highly conserved (***Figure 1A***). This Ser-Thr pair is H-bonded to Coh in binding mode A (***Figure 1A***, black dots). Analogously, binding mode B refers to the configuration where the Ser-Thr pair from helix 3 dominates the H-bonding to Coh (***Figure 1A***, gray dots). Binding mode B was previously crystallized for a homologous *Ct*-Doc (***Carvalho et al., 2003***). Mutation of the Ser-Thr pair in helix 3 to Ala-Ala was used to bias binding and thereby crystallize binding mode A for the same Doc (***Carvalho et al., 2007***). A similar targeted mutagenesis approach was also used to obtain crystal structures of a *Clostridium cellulolyticum* Doc in each binding mode (***Pinheiro et al., 2008***).

To preferentially select for a specific binding mode (A or B), we prepared Doc sequences that incorporated 4 amino acid point mutations, referred to as quadruple mutants ('Q'). To design quadruple mutants, we noted that recent structural work reported a set of *Ct*-Doc domains that differ from the canonical duplicated Ser-Thr sequences. These non-canonical Docs were found to exhibit only a single binding mode (***Brás et al., 2012***; ***Pinheiro et al., 2009***). In one of these non-canonical Doc domains, an Asp-Glu pair was found in place of Ser-Thr. Since the Coh surface is negatively charged, we postulated that including Asp-Glu in place of Ser-Thr within one of the F-hands could be used to effectively knock out a given binding mode for our canonical Doc. Additionally, we incorporated double alanine mutations to replace the conserved Lys-18 Arg-19 residues of a given F-hand motif, further destabilizing a targeted binding mode. Q1 refers to a quadruple mutant where helix 1 has been modified at four positions (i.e. S11D-T12E-K18A-R19A). Q3 refers to the quadruple mutant where helix 3 has been modified at four positions (i.e. S43D-T44E-K50A-R51A). As a negative

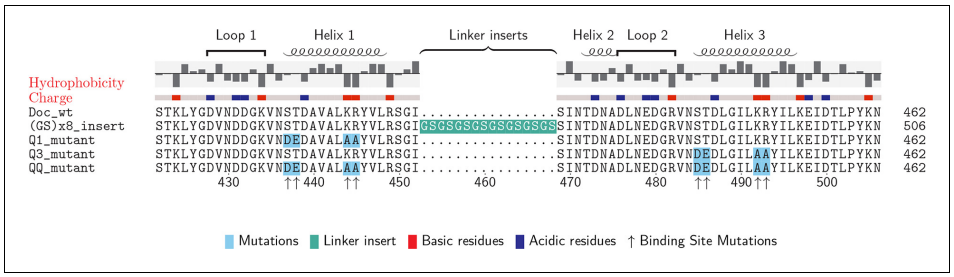


Figure 2. Doc sequences used in this study (N- to C-terminus). Doc_wt: wild-type sequence; hydrophobicity and charge graphs are displayed for the wild-type-Doc (red: positively charged, blue: negatively charged); (GS)x8_insert: A $(Gly\text{-}Ser)_8$ linker was incorporated between helix 1 and helix 2; Q1_mutant: Quadruple mutant in helix 1. Four point mutations (DE/AA) were incorporated into Doc helix 1 to knock out binding mode A; Q3_mutant: Quadruple mutant in helix 3. Four point mutations (DE/AA) were incorporated into Doc helix 3 to knock out binding mode B; QQ_mutant: Non-binding control with both binding modes knocked out. Numbers below indicate amino acid number of the fusion protein construct starting from the xylanase N-terminus.

DOI: 10.7554/eLife.10319.004

control, we prepared a mutant referred to as 'QQ' that incorporated quadruple mutations into both helices 1 and 3.

Doc domains were expressed as fusion domains attached to the C-terminal end of xylanaseT6 (Xyn) from *Geobacillus stearothermophilus* to improve solubility and expression levels as previously reported (***Stahl et al., 2012***). The Xyn domain also acts as a so-called fingerprint in AFM force extension traces to provide a means for screening datasets and searching for known contour length increments. We use the term 'contour length' to refer to the maximum length of a stretched (unfolded) polypeptide chain. Our screening process identified single-molecule interactions and ensured correct pulling geometry. For the Coh domain, we chose cohesin 2 from *Ct*-CipA expressed as a C-terminal fusion domain with the family 3a carbohydrate binding module (CBM) from *Ct*-CipA. In order to exclude artifacts arising from fingerprint domains, protein immobilization or pulling geometry, a second set of fusion proteins was cloned, expressed and probed in complementary experiments using a flavoprotein domain from the plant blue light receptor phototropin (iLOV) (***Chapman et al., 2008***). All protein sequences are provided in the 'Materials and methods' section.

Single-molecule unfolding patterns

The pulling configuration for single-molecule AFM experiments is shown in ***Figure 3A***. CBM-Coh was site-specifically and covalently attached to an AFM cantilever tip and brought into contact with a glass surface modified with Xyn-Doc. The mechanical strength of protein domains and complexes will strongly depend on the pulling points (i.e. sites at which the molecule is attached to cantilever/surface). The site-specific attachment chemistry used here was precisely defined by the chosen residue of immobilization, ensuring the same loading geometry was used on the complex for each and every data trace. After formation of the Coh:Doc complex, the cantilever was retracted at a constant speed that ranged from 200 to 3200 nm/s while the force was monitored by optical cantilever deflection. The resulting force-distance traces were characteristic of the series of energy barriers crossed by the protein complex along the unfolding/unbinding pathway. A sawtooth pattern was consistently observed when molecular ligand-receptor complexes had formed. Sorting the data using contour length transformation (***Puchner et al., 2008***) and identifying traces that contained a Xyn contour length increment (~89 nm) allowed us to screen for single-molecule interactions (***Stahl et al., 2012***), as described in our prior work on Coh:Doc dissociation under force (***Stahl et al., 2012***; ***Schoeler et al., 2014***; ***Jobst et al., 2013***; ***Otten et al., 2014***; ***Schoeler et al., 2015***).

Typical single-molecule interaction traces from such an experiment are shown in ***Figure 3B, C*** and in ***Figure 3—figure supplement 1***. Following PEG linker stretching, an initial set of peaks

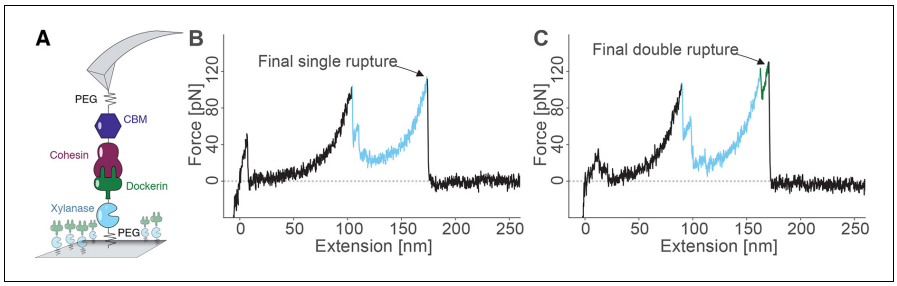


Figure 3. Overview of the experimental configuration and recorded single-molecule unfolding and unbinding traces. (**A**) Schematic depiction showing the pulling geometry with CBM-Coh on the AFM Cantilever and Xyn-Doc on the glass substrate. Each fusion protein is site-specifically and covalently immobilized on a PEG-coated surface. (**B-C**) Each force vs. extension trace shows PEG linker stretching (black), xylanase unfolding and subsequent stretching (blue), and Coh:Doc complex rupture. The Coh:Doc complex rupture occurred in two distinct event types: single (**B**) and double (**C**) ruptures. The 8-nm contour length increment separating the double peaks was assigned to Doc unfolding (C, green).

DOI: 10.7554/eLife.10319.005

The following figure supplement is available for figure 3:

Figure supplement 1. Representative sample of force traces.

DOI: 10.7554/eLife.10319.006

sequentially decreasing in force was assigned to xylanase unfolding and stretching. This domain when unfolded added ~89 nm of free contour length to the system. The final peak (s) corresponded to rupture of the Coh:Doc complex, and occurred as either 'single' or 'double' rupture events. The contour length increment between the two double event peaks was found to be ~8 nm, that is, 8 nm of hidden contour length was added to the biopolymer during a sub-step of Doc unbinding (see 'Discussion'). The 8-nm contour length increment was also observed in complementary experiments employing other fusion domains: xylanase was swapped for an sfGFP domain and CBM was swapped out for an iLOV domain. In these new fusions, the 8 nm Doc increment was still observed, indicating it was not caused by a specific fusion domain. As we show below, double and single rupture events were associated with binding modes A and B, respectively. CBM unfolding length increments (~57 nm) were only rarely observed because the Coh:Doc complex only rarely withstood forces sufficiently high to unfold CBM (***Stahl et al., 2012***).

Ensemble average binding experiments

Binding experiments were carried out in bulk to evaluate the binding affinity of wild-type, Q1, Q3, and QQ Doc sequences to wild-type Coh. Xyn-Doc fusion protein variants were immobilized in a microwell plate and exposed to tag red fluorescent protein (TagRFP) (***Merzlyak et al., 2007***) fused to Coh (TagRFP-Coh) across a range of concentrations, followed by rinsing and subsequent fluorescence readout (***Figure 4A***). The data clearly showed that Q1 and Q3 Doc sequences, each with a mutated binding mode, maintained high-binding affinity with dissociation constants (K_d) in the nM range. These values are in good agreement with previous reports on homologous type-I Doc domains (***Brás et al., 2012***; ***Sakka et al., 2011***). This suggested that mutant Doc domains with one destabilized binding mode were still able to recognize fluorescent protein fused Coh with strong affinity by relying on the alternative binding mode that was preserved. The QQ double knockout mutant, however, showed no appreciable binding over the concentration range tested. This negative control showed that DEAA quadruple mutations were in fact effective at eliminating binding for the targeted modes.

Single-molecule rupture statistics of binding mode mutants

For each Doc tested, we collected tens of thousands of force-extension traces and selected for further analysis only those traces showing the ~89 nm xylanase contour length increments and no other anomalous behavior, resulting in typically 200–3000 usable single-molecule interaction curves per experiment. We determined the number of Coh:Doc unbinding events that occurred as single or double rupture peaks. The results are shown in ***Figure 4B***. The wild-type Doc showed double rupture events in ~57% of the cases, and single rupture events in ~43% of the cases. The mutant designed to knock out binding mode A (Q1), showed a single event probability of ~77%, and a double event probability of ~23%. The mutant designed to knock out binding mode B (Q3) showed a single event probability of ~41%, and a double event probability of ~59%. It is clear from these data that the Q1 mutant has a strong bias toward single peaks that is not observed in the wild-type leading to preliminary assignment of single peaks to binding mode B.

For all double events, we determined the force difference of the second peak relative to the first (***Figure 4C***). Q1 and wild-type on average showed second peaks that were ~15–20% higher in force than the first peak. Q3 meanwhile showed clearly different behavior. Although the ratios of single to double peaks were nearly identical between wild-type and Q3, differences in the relative force between the first and second peaks differentiated wild-type and Q3 (***Figure 4C***). Double peaks for the Q3 mutant were more likely to show a shielded behavior, where the second peak was lower in force than the first peak by ~10%. Although the Q3 mutant showed the same single vs. double event probability as wild-type, the double events for Q3 were distinguishable from those of the wild-type based on this observed decrease in the rupture force of the second peak. The second barrier of the double events was therefore weaker in Q3 than for wild-type. This weaker 2nd double peak for the Q3 mutant combined with similar single/double peak ratios as wild-type leads us to believe that the number of double peaks is being underestimated systematically for the Q3 mutant. Generally, each binding mode still allows for the occurrence of a single event (albeit with different likelihood), in which the whole Doc domain unbinds without an additional unfolding substep. Since the second and final energy barrier for complex dissociation is weaker than the first for the Q3 mutant, the

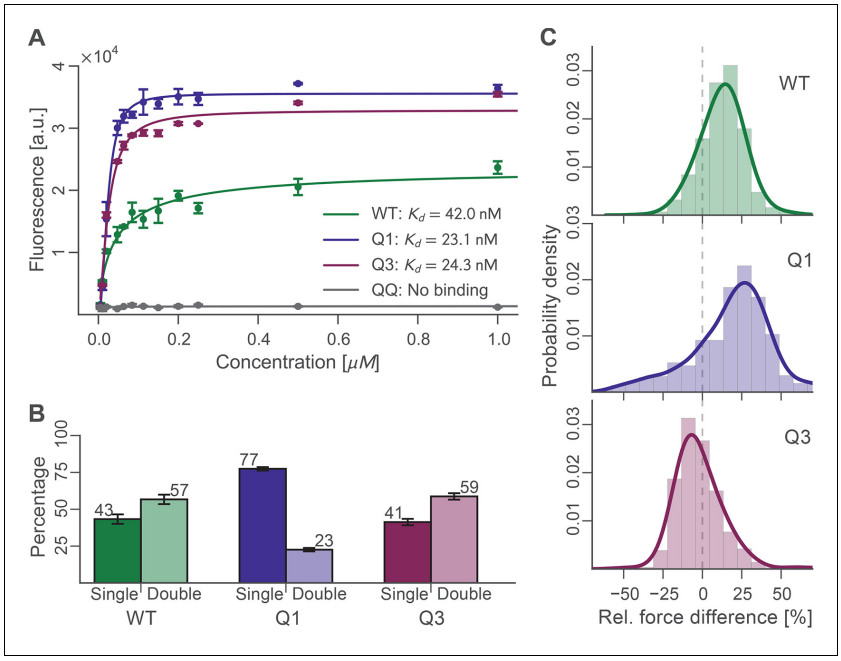


Figure 4. Bulk and single-molecule characterization of Doc mutants. (**A**) Fluorescence binding curve showing binding of TagRFP-labelled Coh to wild-type and mutant Doc nonspecifically immobilized in a 96-well plate. Both Q1 and Q3 mutants bound TagRFP-Coh similarly to wild-type with dissociation constants (K_D) in the low nM range. The negative control QQ mutant showed no binding. Solid lines are 4 parameter logistic nonlinear regression model fits to the data. Error bars represent the standard deviation of three independent samples. (**B**) Event probabilities for single (opaque colors) and double (translucent colors) Coh:Doc rupture peaks determined for Doc wild-type and DE/AA quadruple mutants. Data originate from 947, 4959, and 1998 force-extension traces from wild-type, Q1 and Q3 variants, respectively. Error bars represent 95% Clopper-Pearson confidence intervals based on the beta probability distribution. (**C**) Relative difference in double peak rupture forces for the different Doc variants. Positive values indicate a stronger final peak. Histograms represent concatenated data from various pulling speeds. Drawn lines are kernel density estimates calculated on the raw data.

DOI: 10.7554/eLife.10319.007

The following source data is available for figure 4:

Source data 1. Probability Data.

DOI: 10.7554/eLife.10319.008

probability for the molecule to pass both barriers simultaneously is increased, thus resulting in a higher percentage of single events.

Probing the 8-nm length increment with poly (GS) inserts

We sought to identify the molecular origin of the 8 nm contour length increment separating the double event peaks by engineering additional amino acid sequence length into the Doc domain. Amino acid insert sequences have previously been used to probe length increments in AFM force spectroscopy experiments (***Bertz and Rief, 2009***) (***Carrion-Vazquez et al., 1999***). By adding additional amino acids to the polypeptide chain at a particular location, insert sequences increase the gain in contour length following unfolding of a subdomain in a predictable way. Any change in the observed length increment can be pinpointed to the position in the molecule where the unfolding event occurs. In this case, we engineered flexible $(GS)_8$ insert sequences directly into wild-type Doc between helices 1 and 2, in a flexible loop that was not expected to interfere with either of the two binding modes. Structural homology models (***Figure 5A***) of the wild-type Doc and $(GS)_8$ insert

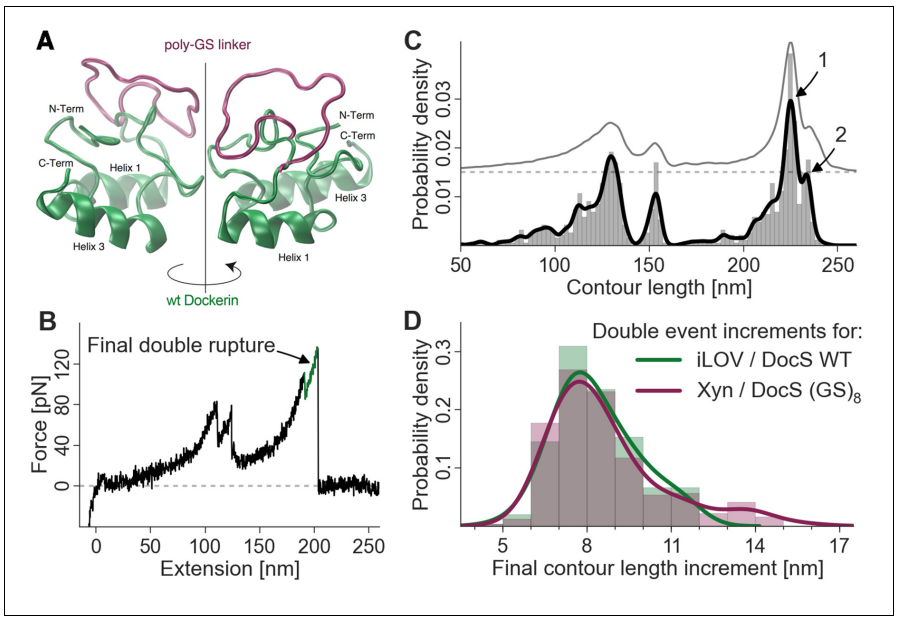


Figure 5. Probing the final contour length increment with Poly (GS) inserts. (**A**) Structural homology model overlay of wild-type and mutant Doc containing a $(GS)_8$-linker between helix 1 and helix 3. The wild-type Doc is shown in green. The 16 amino acid long GS-insert is shown in purple (***Kelley and Sternberg, 2009***) (remaining Doc domain not shown). (**B**) Typical force extension trace with final double rupture event depicted in green (arrow). (**C**) Histogram and kernel density estimate of the transformation of the single force extension trace in panel B into contour length space (black) and kernel density estimate of the whole dataset of single molecule Xyn-Doc:Coh-CBM traces bearing xylanase fingerprint and final double rupture (gray, offset in y-direction for readability) in contour length space. (**D**) Histograms (bars, bin width: 1 nm), kernel density estimates (drawn lines, bandwidth: 0.75 nm, gaussian kernel), and statistical test (Kolmogorov-Smirnov, 'KS test') are each calculated on the raw data of the final increments (peak-to-peak distances) in contour length space (x-distance between arrow 1 and 2 in panel (**C**). Maxima for final double event increments lie at 7.75 nm and 7.73 nm for iLOV-Coh:Doc (wild-type)-sfGFP (N = 255) and Xyn-Doc $(GS)_8$:Coh-CBM (N = 320) final ruptures, respectively (a two-sample KS test on the raw data indicates no significant difference in the data distributions (p-value of 21.7%).

DOI: 10.7554/eLife.10319.009

sequence were calculated using the Phyre server (***Kelley and Sternberg, 2009***). If the 8-nm contour length increment was caused by sequential unbinding of Doc helices 1 and 3 in wild-type Doc, then double peaks for the poly (GS) constructs should show an increase in the double peak contour length increment. As shown in ***Figure 5B,C and D***, the contour length histogram for $(GS)_8$ Doc was indistinguishable from the wild-type Doc. No additional contour length was gained due to additional amino acids inserted between Doc helices 1 and 2. Since the Doc was anchored to the glass slide through an N-terminal xylanase domain, this result indicated that the unfolding event responsible for the 8-nm length increment must be located upstream (i.e. N-terminal) from the site of the $(GS)_8$-insert. This result suggested that unfolding of calcium binding loop 1 and helix 1 in Doc was the source of the 8-nm length increment.

Single-molecule evidence of dual binding mode

To finally confirm the presence of both bound conformations in wild-type Coh:Doc complexes, we replaced xylanase with sfGFPand CBM with iLOV as the contour length marker or fingerprint domains. iLOV was chosen as a superior unfolding fingerprint domain because it does not show multiple unfolding substeps (in contrast to xylanase), which simplified analysis. Also iLOV has an unfolding force distribution that lies in a similar range as the Coh:Doc complex dissociation single and

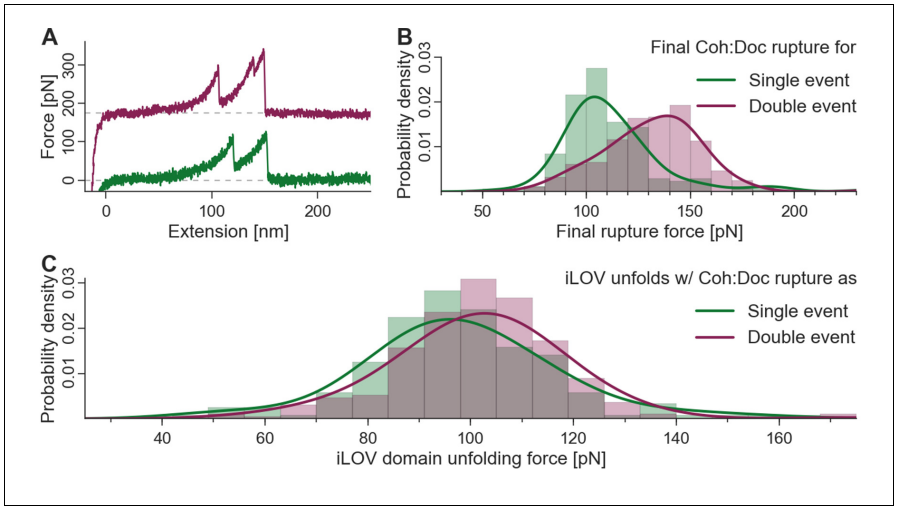


Figure 6. Biasing of unfolding force distributions by dual binding mode. (**A**) Typical force traces showing iLOV unfolding with final single (green) and double (purple) complex ruptures. The curve terminating in a double peak is offset in the y-direction for clarity. (**B**) Final complex rupture force distribution for single and double events. Double events are more mechanically stable. (**C**) iLOV domain unfolding forces for final single (green) and double (red) events at a pulling velocity of 800 nm/s. Histograms (bars), kernel density estimates (lines), and statistical tests are each obtained from the raw data. Maxima for iLOV unfolding lie at 96.0 pN and 102.7 pN for single (N = 172) and double (N = 277) final ruptures, respectively. A two-sample Kolmogorov-Smirnov test showed significant differences in the data distributions (p-value of 0.09%). Since the data were all recorded with a single cantilever and both event types were distributed equally throughout the runtime of the measurement, no systematic biasing is expected. Because of the lower force distribution of final single peaks, the iLOV unfolding force distribution is truncated compared to final double peak force traces, supporting the notion that the binding mode is set prior to mechanical loading of the complex.

DOI: 10.7554/eLife.10319.010

double peaks, allowing for effective biasing of the iLOV unfolding force distributions by the inherent stability difference between single and double event peaks. ***Figure 6A*** shows characteristic single and double event curves containing iLOV unfolding (36-nm contour length increment) followed by Coh:Doc rupture as a single or double event. The rupture force distributions of the single and double event (second peak) ruptures are shown in ***Figure 6B***. The most probable rupture force for single events was ~104 pN, while for double events this value was ~140 pN at a pulling speed of 800 nm/s. We next calculated the unfolding force distributions of the iLOV domain for curves that terminated with single events or double events. If the Coh:Doc complex ruptured before iLOV unfolding was observed, the curve was eliminated from the dataset because it lacked a fingerprint domain length increment. This criterion for inclusion in the dataset results in a biasing of the iLOV unfolding forces, since the maximum of the fingerprint unfolding force distribution that can be observed must lie below that of the Coh:Doc complex. The fact that we observed a downward shift in the iLOV unfolding forces (***Figure 6C***) for curves that terminated in the less mechanically stable single rupture event is confirmation that the single- and double-event peaks arise from separate bound conformations. Each mode has a distinct mechanical stability and energy landscape that is set at the time of receptor-ligand binding, that is once bound, the conformation of the complex does not change. If single- and double-event unbinding patterns were simply two competing pathways out of the same bound state, then the downward shift in rupture force distribution would not be observed for the iLOV unfolding forces. Although this shift in rupture force distributions is comparatively subtle, it can be observed accurately with high statistical significance. We note that the datasets for both binding modes were measured with the same cantilever throughout the runtime of the whole experiment. Calibration and drift issues therefore did not interfere with the required accuracy.

Discussion

The relatively small ~8 kDa Doc domains exhibit an internal sequence and structural symmetry that is believed to give rise to a dual mode of binding to Coh, as shown in ***Figure 1***. In order to study this remarkable plasticity in molecular recognition in greater detail, we prepared a series of mutants (***Figure 2***) designed to either knock out a specific binding mode or add length to the molecule at a specific position. Bulk experiments showed that Doc mutants Q1 and Q3, originally designed to suppress one of the binding modes, were still able to bind Coh with high affinity, while the double knockout did not bind (***Figure 4A***). The equilibrium affinities of Coh binding to Q1, Q3, or wild-type were all similarly high with K_Ds in the low nM range, in good agreement with literature values (***Sakka et al., 2011***), suggesting the two binding modes are thermodynamically equivalent and rendering them indistinguishable with conventional methods such as ELISA or calorimetry. Techniques like surface plasmon resonance could possibly show differing values for on- and off-rates for the mutants, but would still not be able to resolve the binding modes within a wild-type population.

Force spectroscopy with the AFM interrogates individual molecules, and measures their mechanical response to applied force. Since the technique is able to probe individual members of an ensemble, it provided a means to quantify binding mode configurations by assigning unfolding/unbinding patterns to the binding mode adopted by the individual complexes. Site-directed Q1 and Q3 mutations supported the assignment of binding mode A to a characteristic double rupture peak dissociation pathway. Single events were assigned to binding mode B and showed no Doc unfolding substep prior to complex rupture.

We consistently observed 8 nm of added contour length that separated the Doc double peaks. Since force is applied to Doc from the N-terminus, we analyzed the Doc sequence starting at the N-terminus and searched for reasonable portions of Doc that could unfold in a coordinated fashion to provide 8 nm of contour length. The results from the GS-insert experiments (***Figure 5***) indicated no change in the double-event contour length increment, regardless of the added GS-insert length located between helix 1 and 3 in Doc. This result is consistent with the 8 nm length increment being located N-terminally from the GS-insert site, implicating unfolding of Doc calcium binding loop 1 and helix 1 as the source of the 8 nm. This length accurately matches the estimated length increment for unfolding calculated from the crystal structure (***Figure 1D***).

Although this result could also be consistent with the 8 nm increment being located somewhere outside the Doc domain in the polyprotein, we deem this scenario highly unlikely. The 8 nm increment cannot be located in the Xyn or CBM domains because we have accounted for Xyn and CBM lengths in their entirety based on the observed 89 nm and rare 57 nm length increments here and in a previous study (***Stahl et al., 2012***), and for confirmation swapped out those domains for different proteins completely (i.e. iLOV and GFP). The remaining possibility that the 8 nm is located within the Coh domain is also not likely since the barrel-like structure of the Coh is known to be mechanically highly stable (***Valbuena et al., 2009***; ***Hoffmann et al., 2013***). Also, if the 8-nm length increment were due to partial Coh unfolding, the Q1 and Q3 mutants would not be expected to affect the single/double peak ratio or force differences between the double event peaks as was observed (***Figure 4B, C***). The GS-insert data suggest the 8-nm length increment is located within Doc, upstream (N-terminal) from the GS-insert site implicating calcium loop 1 and helix 1 in this unfolding event.

Finally, we observed that an inherent difference in the mechanical stability of single and double event rupture peaks (***Figure 6B***) could be used as a feature by which to discriminate the binding modes. Our analysis algorithm accepted only the force curves that first showed iLOV fingerprint domain unfolding followed by either a single- or double-rupture peak. By observing a small but significant downward shift in the iLOV unfolding force distribution when analyzing curves that terminated in the less stable single-event peak, we confirmed the single-event peaks originate from a unique conformation that is 'set' at the time of complex formation.

Taken together, we propose an unbinding mechanism where the first barrier of the double peaks represents unfolding of the N-terminal calcium binding loop and unraveling of alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19 in the wild-type structure in binding mode A. Based on a length per stretched amino acid of 0.4 nm, the expected contour length for unfolding the Doc domain up to this position would be 7.6 nm, in good agreement with the measured value of 8 nm within experimental error. A portion of the N-terminal calcium binding loop (i.e. residues S11-

T12) is involved in binding to D39 in Coh. The first peak of the double events is attributed to breakage of this interaction and simultaneous unfolding of calcium loop 1 and alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19. Another contributing factor is the intramolecular clasp that has been identified as a stabilizing mechanism among similar type-I Doc domains (**Slutzki et al., 2013**). A recent NMR structural study (**Chen et al., 2014**) on the same wild-type Doc used in this work confirmed a hydrophobic ring-stacking interaction between Tyr-5 and Pro-66. Confirmation of this clasp motif by NMR means the head and tail of the Doc are bound together, additionally stabilizing the barrier that is overcome in the first of the double event peaks. In this scenario, subsequent to breaking the interactions between the calcium binding loop and Coh, disrupting the intramolecular clasp and unfolding the N-terminal loop-helix motif, the remaining bound residues including Lys-18, Arg-19, Lys-50, Leu-54, and Lys-55 stay bound to Coh and are able to withstand substantial force on their own, eventually breaking in the second and final of the double rupture peaks. Prior work further supports this unbinding mechanism, revealing that a progressive N-terminal truncation of Doc did not affect the interaction largely, unless the truncation reached the Lys-18 and Arg-19 residues (**Karpol et al., 2009**). This corroborates the idea of the C-terminal end of helix 1 being a crucial part of the binding site within the complex. Single rupture peaks were thus observed when the wild-type complex was bound in binding mode B, and no unfolding of Ca-binding loop 1 or helix 1 occurred. Force was propagated directly to bound residues Lys-18, Leu-22, and Arg-23 which when broken resulted in complete complex dissociation.

Given the fingerprint biasing phenomenon (**Figure 6C**), we finally sought to correct the single/double peak counting statistics (**Figure 4B**) in order to correct for undercounting of single peaks due solely to their failure to reach sufficiently high forces to unfold the fingerprint domain. Only traces showing a fingerprint were analyzed to ensure defined unfolding geometry. Using the rupture force distributions of singles, doubles, iLOV, and xylanase domains, we calculated the probability of occurrence of fingerprint unfolding at a force higher than the single-event ruptures. This overlap probability was found to be 0.85 for iLOV and 0.40 for xylanase. When the single/double peak ratios for were corrected for this effect, the final binding mode ratios (binding mode A/binding mode B, i.e., doubles/singles) were found to be 0.95 and 0.87 for xylanase-Doc and iLOV, respectively. These ratios are close to 1 indicating comparable probability of each binding mode after accounting for biasing the single/double peak counting statistics due to fingerprint domain stability. We note that these numbers are also slightly lower than unity due to the exclusion of double peaks that occurred before unfolding of the fingerprint domains. Further details on rupture force distributions and overlap statistics are shown in **Figure 7**. As the magnitude of biasing changes with the unfolding force distributions of each fingerprint domain, overlaps in the probability distributions allow for normalizing single/double event ratios of experimental data sets with different fingerprinting domains. For the Coh:Doc complex unbinding event, biasing (undercounting) is more pronounced for the mechanically weaker single ruptures. This normalization procedure shows the relative difference of biasing between single and double events, as double events are less biased than single events.

The biological significance of Coh-Doc interactions in the context of cellulosome assembly and catalysis cannot be overstated. Their high affinity and specificity, along with their modularity, thermostability, and their ultrastable mechanical properties all make Coh-Doc unique from a biophysics perspective, and attractive from an engineering standpoint. Dual binding mode Doc domains are broadly predicted among many cellulosome producing bacteria (e.g. *C. thermocellum, C. cellulolyticum, R. flavefaciens*), however relatively few have been confirmed experimentally (**Carvalho et al., 2007**; **Pinheiro et al., 2008**; **Brás et al., 2012**). In fact, the direct effect of single vs. dual binding modes on the ability of cellulosomes to convert substrate into sugars is currently unknown. It is therefore unclear whether or not dual binding modes affect, for example, the catalytic properties of native or engineered synthetic cellulosomes.

However, it is important to note that cellulosome producing bacteria invariably live among communities with other microorganisms, which may be producing cellulases and cellulosomes of their own. In such an environment, a dual binding mode could enable organisms to produce enzymes that are able to bind to a neighboring species' scaffoldins, yet still retain high-affinity interactions with host scaffoldins. They would be able to combine resources with neighboring cells in a mixed microbial consortium. The dual binding mode could therefore allow genetic drift to explore interspecies protein binding. Indeed, cross-species reactivity between Coh and Doc has been reported (**Haimovitz et al., 2008**). Cellulosome-producing microbes may therefore be pursuing a middle

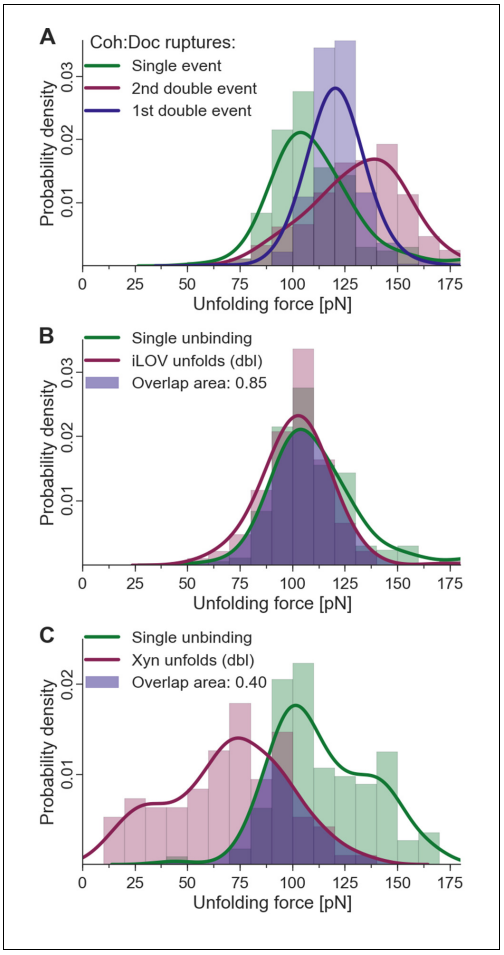


Figure 7. Fingerprint unfolding and complex unbinding forces. (**A**) Rupture force distribution of final complex ruptures for single (green), first (purple) and second (red) double unbinding events. (**B**) Overlap area (purple) of iLOV domain unfolding force distribution (red) (iLOV-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. (**C**) Overlap area (purple) of Xyn domain unfolding force distribution (red) (Xyn-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. Overlaps in probability distributions allow normalizing single-event counts to double events to account for different biasing caused by the different unfolding forces of the fingerprint domain. Biasing occurs, because for overlapping force distributions of fingerprint unfolding and complex ruptures, unbinding events are more likely to take place without fingerprint unfolding if the two distributions are closer together. For the Coh:Doc unbinding, this effect is more pronounced for the weaker single ruptures. Because double events are also biased, this still does not give a true quantification, but only compensates for the differences of biasing. The non-bell-evans-like shape of the single rupture peaks, especially in the region of the 1st double event peak (**A**) suggests that this class of curves does not contain a single type of unbinding mechanism, but rather a superposition of different event types.

DOI: 10.7554/eLife.10319.011

ground between protein synthesis strictly for selfish vs. communal usage. By distinguishing the presence of each binding mode for wild-type Doc domains, the single-molecule biophysical approach presented here based on differences in mechanical hierarchies will facilitate further study into the significance of the dual binding mode.

In summary, the dual binding mode of Coh:Doc domains has so far proven resistant to explicit experimental characterization. Crystallography combined with mutagenesis has provided snapshots of the two modes, but resolving each of the modes for wild-type Doc under near native conditions has up until now not been possible. We have demonstrated the advantages of a single-molecule approach in resolving these subtle differences in molecular conformations of bound complexes. Despite having equal thermodynamic binding affinity, when mechanically dissociated by pulling from the N-terminus of Doc, binding mode A was more mechanically stable with an additional energy barrier. This mechanical difference was exploited to probe the two binding modes independently from one another, providing direct observation of this unique mechanism in molecular recognition. In the future, harnessing control over binding modes could offer new approaches to designing molecular assembly systems that achieve defined protein orientations.

Materials and methods

Site-directed mutagenesis of plasmid DNA

A pET28a vector containing the previously cloned xylanaseT6 from *Geobacillus stearothermophilus* (***Salama-Alber et al., 2013***) and DocS dockerin from *Clostridium thermocellum* Cel48S were subjected to QuikChange mutagenesis (***Wang and Malcolm, 1999***) to install the following mutations: Q1, Q3, and QQ in the dockerin and T129C in the xylanase, respectively.

For insertion of the $(GS)_4$ and $(GS)_8$ linkers into the Doc domain, exponential amplification with primers bearing coding sequences for the inserts at their 5′-ends was performed with a Phusion High-Fidelity DNA polymerase (New England Biolabs, MA). PCR products were then blunt end ligated using KLD Enzyme Mix and KLD Reaction Buffer from the Q5 site directed mutagenesis kit (New England Biolabs, MA). The modified DNA constructs were used to transform *Escherichia coli* DH5-alpha cells, grown on kanamycin-containing agar plates and subsequently screened. All mutagenesis products were confirmed by DNA sequencing analysis.

Primers used for inserting the $(GS)_8$ linker into the Doc domain:

Fw 5′-ggttctggctccggttctggctccagcatcaacactgacaat-3′

Rev 5′-agaaccggagccagagccggaacctatacctgatctcaaaacatatct-3′

Protein expression and purification

Fusion proteins HIS-CBM A2C-Coh2 (*C.t.*) were expressed in *E. coli* BL21(DE3)RIPL cells in kanamycin-containing media supplemented with 2mM calcium chloride overnight at 16°C. After harvesting, cells were lysed by sonication, and the lysate was subjected to heat treatment at 60°C for 30 min to precipitate the bulk of the host bacterial proteins, leaving the expressed thermophilic proteins in solution. The lysate was then pelleted, and the supernatant fluids were applied to a beaded cellulose column and incubated at 4°C for 1 hr. The column was then washed with 50 mM Tris buffer (pH 7.4) containing 1.15 M NaCl, and the protein was eluted using a 1% (vl/v) triethylamine aqueous solution. Tris buffer was added to the eluent and the solution was neutralized with HCl.

Fusion proteins HIS-Xyn T129C-DocS (*C.t.*) wild-type, Q1, and Q3 mutants were expressed as described above. Following heat treatment, the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20mM imidazole and 2mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was then dialyzed to remove the imidazole.

Fusion proteins ybbR-HIS-CBM A2C-Coh2 (*C.t.*), ybbR-HIS-Xyn T129C-DocS (*C.t.*) wild-type and QQ mutants and ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_4$ insert were expressed in *E. coli* BL21(DE3) RIPL cells; ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_8$ insert fusion protein variants were expressed in *E. coli* NiCo21(DE3)RIPL cells. Cultivation and expression was done in ZYM-5052 autoinduction media (***Studier, 2005***) containing kanamycin (and chloramphenicol, in case of the NiCo21(DE3)RIPL cells) overnight at 22°C, overall 24 hr. After harvesting, cells were lysed using sonication. The lysate was then pelleted by centrifugation at 39,000 rcf, the supernatant fluids were applied to Ni-NTA columns

and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with polyacrylamide gravity flow columns or with polyacrylamide spin desalting columns.

All protein solutions were concentrated with Amicon centrifugal filter devices and stored in 50% (v/v) glycerol at -20°C (ybbR-free constructs) or -80°C (ybbR-bearing constructs). The concentrations of the protein stock solutions were determined to be in the order of 1–15 mg/mL by absorption spectrophotometry at a wavelength of 280 nm.

ELISA-like binding assay

1 μM of Xyn-Doc fusion proteins (wild-type Q1, Q3, QQ Doc fusions) bearing either wild-type or mutant Doc domains were adsorbed onto surfaces of the wells of a 96-well nunc maxi sorp plate (Thermo Scientific, Pittsburgh, PA). After blocking (2% (w/v) BSA, 0.05% Tween 20 in TBS buffer) and several rinsing steps, a red fluorescent protein-cohesin (StrepII-TagRFP-Coh2 (*C.t.*), Addgene ID 58,710 (***Otten et al., 2014***)) fusion construct was incubated to the unspecifically immobilized Doc fusion proteins over a range of concentrations. After further rinsing, the fluorescence of the TagRFP domain was measured with a multi-well fluorescence plate reader (M1000 PRO, Tecan Group Ltd., Männedorf, Switzerland). Fluorescence values were plotted against their corresponding concentration values for each protein variant, and 4 parameter logistic nonlinear regression model functions were fitted to the data to determine the transition point of the curve.

Surface immobilization strategies

The Xyn domain had a cysteine point mutation at position 129 (Xyn T129C) to facilitate covalent attachment to a glass surface *via* Polyethylene glycol (PEG)-maleimide linkers. There were no other cysteines within the Xyn or Doc domains, which ensured site-specific immobilization of the molecule and defined mechanical loading of Doc from the N-terminus for the AFM experiments. The CBM domain likewise contained an A2C cysteine point mutation for covalent attachment to the cantilever tip *via* PEG-maleimide linkers. The second set of fusion proteins sfGFP-Doc and iLOV-Coh was covalently attached to coenzyme A bearing PEG linkers by their terminal ybbR tags.

AFM sample preparation

For AFM measurements, silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation nominal spring constant: 100 pN/nm; 25 kHz resonance frequency in water), and glass coverslips (Menzel Gläser, Braunschweig, Germany; diameter 22mm) were used. 3-Aminopropyl dimethyl ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany), α-Maleinimidohexanoic-ω-NHS PEG (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany; PEG-MW: 5 kDa), immobilized tris (2-carboxyethyl)phosphine (TCEP) disulfide reducing gel (Thermo Scientific, Pittsburgh, PA), tris (hydroxymethyl) aminomethane (TRIS, >99% p.a., Carl Roth, Karlsruhe, Germany), $CaCl_2$ (>99% p.a., Carl Roth, Karlsruhe, Germany), sodium borate (>99.8% p.a., Carl Roth, Karlsruhe, Germany), NaCl (>99.5% p.a., Carl Roth, Karlsruhe, Germany), ethanol (>99% p.a.), toluene (>99.5% p.a., Carl Roth, Karlsruhe, Germany) were used as received. Sodium borate buffer was 150 mM, pH 8.5. Measurement buffer for AFM-SMFS was tris-buffered saline supplemented with 1 mM $CaCl_2$ (TBS, 25 mM TRIS, 75 mM NaCl, 1 mM $CaCl_2$ pH 7.2). All buffers were filtered through a sterile 0.2 μm polyethersulfone membrane filter (Nalgene, Rochester, NY) prior to use.

Force spectroscopy measurement samples, measurements and data analysis were prepared and performed according to previously published protocols (***Jobst et al., 2013***;***Otten et al., 2014***). In brief, NHS-PEG-Maleimide linkers were covalently attached to cleaned and amino-silanized silicon nitride AFM cantilevers and cover glasses. The respective protein constructs were covalently linked either *via* engineered cysteine residues to the maleimide groups of the surface on the sample directly, or *via* Sfp phosphopantetheinyl transferase-mediated attachment of a terminal ybbR tag to coenzyme A, which was previously attached to the maleimide groups of the surface.

AFM-SMFS measurements

AFM data were recorded in 25 mM TRIS pH 7.2, 75 mM NaCl and 1mM $CaCl_2$ buffer solution (TBS). Retraction velocities for constant speed force spectroscopy measurements varied between 0.2 and 3.2 μm/s. Cantilever spring constants were calibrated utilizing the thermal method applying the

equipartition theorem to the one dimensionally oscillating lever (**Hutter and Bechhoefer, 1993**; **Cook et al., 2006**). Measurements were performed on custom built instruments, deploying an Asylum Research (Santa Barbara, CA, USA) MFP-3D AFM controller and Physik Instrumente (Karlsruhe, Germany) or attocube (Munich, Germany) piezo nanopositioners (**Gumpp et al., 2009**). After each measurement, the xy-stage was actuated by 100 nm to probe a new spot on the surface and measure new individual Xyn-Doc fusion molecules. Instrument control software was programmed in Igor Pro 6.3 (Wavemetrics). The retraction speed was controlled with a closed-loop feedback system running internally on the AFM controller field-programmable gate array (FPGA).

Force-extension data analysis

Data analysis and plotting was performed in Python (Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org) utilizing the libraries NumPy and SciPy (**van der Walt et al., 2011**) and Matplotlib (**Hunter, 2007**).

Measured raw data were analyzed by determining the zero force value with the baseline position and applying a cantilever bending correction to the z-position. The resulting force distance traces were coarsely screened for peaks as sudden drops in force and curves with less than three peaks (such as in **Figure 3—figure supplement 1**, panel F) were excluded, as they contain no clearly identifiable signal. Force-distance traces were transformed into contour length space with the inverse worm-like-chain model (**Jobst et al., 2013**), assuming a fixed persistence length of 0.4 nm. Screening for the 89 nm xylanase, the 36nm iLOV and the final 8 nm final double rupture increment was performed by finding their corresponding local maxima in a kernel density estimate with bandwidth b = 1 nm. Thresholds in force, distance, and peak counts were applied to sort out nonspecific and multiple interactions. All curves were ultimately selected for the xylanase or iLOV fingerprint and checked manually. For the counting statistics, double peaks were detected as an increment of 8 +- 4 nm in contour length for final rupture peaks in the contour length plot, given that the curve showed one of the fingerprints. If a double peak was detected, the force difference was determined as the percentual difference between the first and the final rupture peak force.

Barrier position diagrams were assembled using optimal alignment through cross-correlation (**Puchner et al., 2008**; **Otten et al., 2014**). The numbers of points included in fitted histograms are provided in the figure captions, along with the statistical tests and significance values obtained.

Amino acid sequences

pET28a-HIS-XynT129C-DocS (*C.t.*) wild-type

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALK-
RYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-XynT129C-DocS (*C.t.*) 16aa GS Insert

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAV-
EPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVP-
QWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWY-
QIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSE-
AEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFW-
GIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGT-
PSTKLYGDVNDDGKVNSTDAVALKRYVLRSGIGSGSGSGSGSGSGSGSSINTDNADLNEDGRVNSTDLGI-
LKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q1 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-

QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNDEDAVALA-
AYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q3 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALK-
RYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) QQ mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNDEDAVALA-
AYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-ybbR-HIS-sfGFP-DocIS (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEG-
DATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYK-
TRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQL-
ADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKKVVPGTPST-
KLYGDVNDDGKVNSTDAVALKRYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-CBM A2C-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASMCNTVSGNLKVEFYNSNPSDTTNSINPQFKVTNT-
GSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTG-
GTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPSTQP-
VTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEII-
GIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFAD-
NDLVEQKVSFIDGGVNVGNAT

pET28a-ybbR-HIS-iLOV-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASGSPEFIEKNFVITDPRLPDNPIIFASDGFLELTEYSR-
EEILGRNARFLQGPETDQATVQKIRDAIRDQRETTVQLINYTKSGKKFWNLLHLQPVRDQKGELQYFIGV-
QLDGSDHVGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSK-
GIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATV-
KSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

pET28a-StrepII-TagRFP-Coh2 (*C.t.*)

MWSHPQFEKVSKGEELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDI-
LATSFMYGSRTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFPS-
NGPVMQKKTLGWEANTEMLYPADGGLEGRSDMALKLVGGGHLICNFKTTYRSKKPAKNLKMPGVYYVD-
HRLERIKEADKETYVEQHEVAVARYCDLPSKLGHKLNGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSD-
GVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKII-
VFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

Acknowledgements

The authors acknowledge Carlos Fontes, Sarah Teichmann, Stefan Stahl, and Ellis Durner for helpful discussions. Support for this work was provided by the ERC Advanced Grant CelluFuel, and the EU 7th Framework Programme NMP4- SL-2013-604530 (CellulosomePlus), and the German-Israeli Foundation (GIF) for Scientific Research and Development. MAN acknowledges support from Society in Science – The Branco Weiss Fellowship from ETH Zurich.

Additional information

Funding

Funder	Grant reference number	Author
European Research Council	294438	Hermann E Gaub
European Commission	NMP4- SL-2013-604530	Daniel B Fried
German-Israeli Foundation for Scientific Research and Development	G-147-207.4-2012	Edward A Bayer Hermann E Gaub Michael A Nash
Society in Science	Branco Weiss Fellowship	Michael A Nash

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MAJ, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article, Contributed reagents; LFM, Conception and design, Analysis and interpretation of data, Drafting and revising the article, Contributed unpublished essential data or reagents; CS, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article; WO, Acquisition of data, Contributed reagents, Drafting and revising the article; DBF, EAB, Conception and design, Contributed reagents, Drafting and revising the article; HEG, MAN, Conception and design, Analysis and interpretation of data, Drafting and revising the article

References

Bayer EA, Belaich JP, Shoham Y, Lamed R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* **58**:521–554. doi: 10.1146/annurev.micro.57.030502.091022

Beitz E. 2000. TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **16**:135–139. doi: 10.1093/bioinformatics/16.2.135

Bertz M, Rief M. 2009. Ligand binding mechanics of maltose binding protein. *Journal of Molecular Biology* **393**:1097–1105. doi: 10.1016/j.jmb.2009.08.066

Brás JL, Alves VD, Carvalho AL, Najmudin S, Prates JA, Ferreira LM, Bolam DN, Romão MJ, Gilbert HJ, Fontes CM. 2012. Novel clostridium thermocellum type i cohesin-dockerin complexes reveal a single binding mode. *The Journal of Biological Chemistry* **287**:44394–44405. doi: 10.1074/jbc.M112.407700

Béguin P, Aubert J-P. 1994. The biological degradation of cellulose. *FEMS Microbiology Reviews* **13**:25–58. doi: 10.1111/j.1574-6976.1994.tb00033.x

Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM. 1999. Atomic force microscopy captures length phenotypes in single proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**:11288–11292. doi: 10.1073/pnas.96.20.11288

Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA. 2003. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **100**:13809–13814. doi: 10.1073/pnas.1936124100

Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith N, Bayer EA, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA, Gilbert HJ. 2007. Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America* **104**:3089–3094. doi: 10.1073/pnas.0611173104

Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, Oparka KJ, Christie JM. 2008. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20038–20043. doi: 10.1073/pnas.0807551105

Chen C, Cui Z, Xiao Y, Cui Q, Smith SP, Lamed R, Bayer EA, Feng Y, Cui Q, Smith SP, Lamed R, et al. 2014. Revisiting the NMR solution structure of the Cel48S type-I dockerin module from clostridium thermocellum reveals a cohesin-primed conformation. *Journal of Structural Biology* **188**:188–193. doi: 10.1016/j.jsb.2014.09.006

Cook SM, Lang KM, Chynoweth KM, Wigton M, Simmonds RW, Schäffer TE. 2006. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnology* **17**: 2135–2145. doi: 10.1088/0957-4484/17/9/010

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190. doi: 10.1101/gr.849004

Currie MA, Adams JJ, Faucher F, Bayer EA, Jia Z, Smith SP. 2012. Scaffoldin conformation and dynamics revealed by a ternary complex from the clostridium thermocellum cellulosome. *The Journal of Biological Chemistry* **287**:26953–26961. doi: 10.1074/jbc.M112.343897

Demain AL, Newcomb M, Wu JHD. 2005. Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews* **69**:124–154. doi: 10.1128/MMBR.69.1.124-154.2005

Engel A, Müller DJ. 2000. Observing single biomolecules at work with the atomic force microscope. *Nature Structural Biology* **7**:715–718. doi: 10.1038/78929

Fontes CM, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual Review of Biochemistry* **79**:655–681. doi: 10.1146/annurev-biochem-091208-085603

Gumpp H, Stahl SW, Strackharn M, Puchner EM, Gaub HE. 2009. Ultrastable combined atomic force and total internal fluorescence microscope. *Review of Scientific Instruments* **80**:063704. doi: 10.1063/1.3148224

Haimovitz R, Barak Y, Morag E, Voronov-Goldman M, Shoham Y, Lamed R, Bayer EA. 2008. Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8**:968–979. doi: 10.1002/pmic.200700486

Hall BA, Sansom MSP. 2009. Coarse-grained MD simulations and protein−protein interactions: the cohesin−dockerin system. *Journal of Chemical Theory and Computation* **5**:2465–2471. doi: 10.1021/ct900140w

Hoffmann T, Tych KM, Hughes ML, Brockwell DJ, Dougan L. 2013. Towards design principles for determining the mechanical stability of proteins. *Physical Chemistry Chemical Physics* **15**:15767–15780. doi: 10.1039/c3cp52142g

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**:33–38. doi: 10.1016/0263-7855(96)00018-5

Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**:90–95. doi: 10.1109/MCSE.2007.55

Hutter JL, Bechhoefer J. 1993. Calibration of atomic-force microscope tips. *Review of Scientific Instruments* **64**: 1868 doi: 10.1063/1.1143970

Jobst MA, Schoeler C, Malinowska K, Nash MA. 2013. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *Journal of Visualized Experiments* **82** doi: 10.3791/50950

Karpol A, Kantorovich L, Demishtein A, Barak Y, Morag E, Lamed R, Bayer EA. 2009. Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction. *Journal of Molecular Recognition* **22**:91–98. doi: 10.1002/jmr.926

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**:363–371. doi: 10.1038/nprot.2009.2

Kufer SK, Puchner EM, Gumpp H, Liedl T, Gaub HE. 2008. Single-molecule cut-and-paste surface assembly. *Science* **319**:594–596. doi: 10.1126/science.1151424

Kufer SK, Strackharn M, Stahl SW, Gumpp H, Puchner EM, Gaub HE. 2009. Optically monitoring the mechanical assembly of single molecules. *Nature Nanotechnology* **4**:45–49. doi: 10.1038/nnano.2008.333

Li H, Cao Y. 2010. Protein mechanics: from single molecules to functional biomaterials. *Accounts of Chemical Research* **43**:1331–1341. doi: 10.1021/ar100057a

Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TW, Chudakov DM. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* **4**:555–557. doi: 10.1038/nmeth1062

Otten M, Ott W, Jobst MA, Milles LF, Verdorfer T, Pippig DA, Nash MA, Gaub HE, Milles LF, Verdorfer T, Pippig DA, et al. 2014. From genes to protein mechanics on a chip. *Nature Methods* **11**:1127–1130. doi: 10.1038/nmeth.3099

Pagès S, Bélaïch A, Bélaïch JP, Morag E, Lamed R, Shoham Y, Bayer EA. 1997. Species-specificity of the cohesin-dockerin interaction between clostridium thermocellum and clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain. *Proteins* **29**:517–527.

Pinheiro BA, Proctor MR, Martinez-Fleites C, Prates JA, Money VA, Davies GJ, Bayer EA, Fontesm CM, Fierobe HP, Gilbert HJ. 2008. The clostridium cellulolyticum dockerin displays a dual binding mode for its cohesin partner. *The Journal of Biological Chemistry* **283**:18422–18430. doi: 10.1074/jbc.M801533200

Pinheiro BA, Gilbert HJ, Sakka K, Sakka K, Fernandes VO, Prates JA, Alves VD, Bolam DN, Ferreira LM, Fontes CM. 2009. Functional insights into the role of novel type I cohesin and dockerin domains from clostridium thermocellum. *The Biochemical Journal* **424**:375–384. doi: 10.1042/BJ20091152

Puchner EM, Franzen G, Gautel M, Gaub HE. 2008. Comparing proteins by their unfolding pattern. *Biophysical Journal* **95**:426–434. doi: 10.1529/biophysj.108.129999

Roberts E, Eargle J, Wright D, Luthey-Schulten Z. 2006. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**:382 doi: 10.1186/1471-2105-7-382

Sakka K, Sugihara Y, Jindou S, Sakka M, Inagaki M, Sakka K, Kimura T. 2011. Analysis of cohesin-dockerin interactions using mutant dockerin proteins. *FEMS Microbiology Letters* **314**:75–80. doi: 10.1111/j.1574-6968.2010.02146.x

Salama-Alber O, Jobby MK, Chitayat S, Smith SP, White BA, Shimon LJ, Lamed R, Frolow F, Bayer EA. 2013. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *The Journal of Biological Chemistry* **288**:16827–16838. doi: 10.1074/jbc.M113.466672

Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, Ott W, Fried DB, Bayer EA, Schulten K, Gaub HE, Nash MA. 2014. Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications* **5**:5635–35. doi: 10.1038/ncomms6635

Schoeler C, Bernardi RC, Malinowska KH, Durner E, Ott W, Bayer EA, Schulten K, Nash MA, Gaub HE. 2015. Mapping mechanical force propagation through biomolecular complexes. *Nano Letters* **15**:7370–7376. doi: 10.1021/acs.nanolett.5b02727

Schwarz WH. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* **56**:634–649. doi: 10.1007/s002530100710

Slutzki M, Jobby MK, Chitayat S, Karpol A, Dassa B, Barak Y, Lamed R, Smith SP, Bayer EA, Barak Y, et al. 2013. Intramolecular clasp of the cellulosomal ruminococcus flavefaciens ScaA dockerin module confers structural stability. *FEBS Open Bio* **3**:398–405. doi: 10.1016/j.fob.2013.09.006

Smith SP, Bayer EA. 2013. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Current Opinion in Structural Biology* **23**:686–694. doi: 10.1016/j.sbi.2013.09.002

Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, Gaub HE. 2012. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **109**:20431–20436. doi: 10.1073/pnas.1211929109

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**:207–234. doi: 10.1016/j.pep.2005.01.016

Valbuena A, Oroz J, Hervas R, Vera AM, Rodriguez D, Menendez M, Sulkowska JI, Cieplak M, Carrion-Vazquez M. 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences of the United States of America* **106**:13791–13796. doi: 10.1073/pnas.0813093106

Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques* **26**:680–682.

Woodside MT, Block SM. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annual Review of Biophysics* **43**:19–39. doi: 10.1146/annurev-biophys-051013-022754

van der Walt Stefan, Colbert SC, Varoquaux Gael. 2011. The NumPy array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **13**:22–30. doi: 10.1109/MCSE.2011.37

Bibliography

[1] B T Marshall, M Long, J W Piper, T Yago, R P McEver, and C Zhu. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*, 423(6936):190–193, 2003.

[2] Jongseong Kim, Nathan E Hudson, and Timothy A Springer. Force-induced on-rate switching and modulation by mutations in gain-of-function von Willebrand diseases. *Proceedings of the National Academy of Sciences*, page 201501689, March 2015.

[3] Jochen P Mueller, Salome Mielke, Achim Loef, Tobias Obser, Christof Beer, Linda K Bruetzel, Diana A Pippig, Willem Vanderlinden, Jan Lipfert, Reinhard Schneppenheim, and Martin Benoit. Force sensing by the vascular protein von Willebrand factor is tuned by a strong intermonomer interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 113(5):1208–1213, 2016.

[4] O Salama-Alber, M K Jobby, S Chitayat, S P Smith, B A White, L J W Shimon, R Lamed, F Frolow, and E A Bayer. Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *Journal of Biological Chemistry*, April 2013.

[5] Constantin Schoeler, Klara H Malinowska, Rafael C Bernardi, Lukas F Milles, Markus A Jobst, Ellis Durner, Wolfgang Ott, Daniel B Fried, Edward A Bayer, Klaus Schulten, Hermann E Gaub, and Michael A Nash. Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications*, 5:1–8, December 2014.

[6] Edward A Bayer, Ely Morag, and Raphael Lamed. The cellulosome — A treasure-trove for biotechnology. *Trends in Biotechnology*, 12(9):379–386, September 1994.

[7] Wolfgang Ott, Markus A Jobst, Constantin Schoeler, Hermann E Gaub, and Michael A Nash. Single-molecule force spectroscopy on polyproteins and receptor-ligand complexes: the current toolbox. *Journal of Structural Biology*, February 2016.

[8] G Binnig and C F Quate. Atomic Force Microscope. *Physical Review Letters*, 56(9):930–933, March 1986.

[9] B Drake, C Prater, A Weisenhorn, S Gould, T Albrecht, C Quate, D Cannell, H Hansma, and P Hansma. Imaging crystals, polymers, and processes in water with the atomic force microscope. *Science*, 243(4898):1586–1589, March 1989.

[10] D J Müller, G Büldt, and A Engel. Force-induced conformational change of bacteriorhodopsin. *J. Mol. Biol.*, 249(2):239–243, 2 June 1995.

[11] M Radmacher, M Fritz, C M Kacher, J P Cleveland, and P K Hansma. Measuring the viscoelastic properties of human platelets with the atomic force microscope. *Biophys. J.*, 70(1):556–567, January 1996.

[12] M Radmacher, R W Tillamnn, M Fritz, and H E Gaub. From molecules to cells: imaging soft samples with the atomic force microscope. *Science*, 257(5078):1900–1905, 25 September 1992.

[13] S M Block, L S Goldstein, and B J Schnapp. Bead movement by single kinesin molecules studied with optical tweezers. *Nature*, 348(6299):348–352, 22 November 1990.

[14] E L Florin, M Rief, H Lehmann, M Ludwig, C Dornmair, V T Moy, and H E Gaub. Sensing Specific Molecular-Interactions with the Atomic-Force Microscope. *Biosens. Bioelectron.*, 10(9-10):895–901, 1995.

[15] Gil U Lee, David A Kidwell, and Richard J Colton. Sensing discrete Streptavidin-Biotin interactions with atomic force microscopy. *Langmuir*, 10(2):354–357, 1994.

[16] Gil U Lee, Linda A Chrisey, and Richard J Colton. Direct measurement of the forces between complementary strands of DNA. *Science*, 266(5186):771–773, 4 November 1994.

[17] S B Smith, L Finzi, and C Bustamante. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 13 November 1992.

[18] K Svoboda, C F Schmidt, B J Schnapp, and S M Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365(6448):721–727, 21 October 1993.

[19] M Rief, Mathias Gautel, Filipp Oesterhelt, Julio M Fernandez, and Hermann E Gaub. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*, 276 (5315):1109–1112, May 1997.

[20] M Carrion-Vazquez, A F Oberhauser, S B Fowler, P E Marszalek, S E Broedel, J Clarke, and J M Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U. S. A.*, 96(7):3694–3699, 30 March 1999.

[21] Andres F Oberhauser, Piotr E Marszalek, Harold P Erickson, and Julio M Fernandez. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681): 181–185, 14 May 1998.

[22] F Oesterhelt, D Oesterhelt, M Pfeiffer, A Engel, H E Gaub, and Daniel J Müller. Unfolding pathways of individual bacteriorhodopsins. *Science*, 288(5463):143–146, 7 April 2000.

[23] David J Brockwell, Emanuele Paci, Rebecca C Zinober, Godfrey S Beddard, Peter D Olmsted, D Alastair Smith, Richard N Perham, and Sheena E Radford. Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nature Structural Biology*, 10(9): 731–737, August 2003.

[24] Mariano Carrión-Vázquez, Hongbin Li, Hui Lu, Piotr E Marszalek, Andres F Oberhauser, and Julio M Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.*, 10(9):738–743, 2003.

[25] Hendrik Dietz, Felix Berkemeier, Morten Bertz, and Matthias Rief. Anisotropic deformation response of single protein molecules. *Proceedings of the National Academy of Sciences*, 103(34):12724–12728, 2006.

[26] Minkyu Kim, Chien-Chung Wang, Fabrizio Benedetti, Mahir Rabbi, Vann Bennett, and Piotr E Marszalek. Nanomechanics of streptavidin hubs for molecular materials. *Adv. Mater.*, 23(47):5684–5688, December 2011.

[27] Xiaotang Hu and Hongbin Li. Force spectroscopy studies on protein-ligand interactions: a single protein mechanics perspective. *FEBS Lett.*, 588(19):3613–3620, 1 October 2014.

[28] Jonne Helenius, Carl-Philipp Heisenberg, Hermann E Gaub, and Daniel J Müller. Single-cell force spectroscopy. *J. Cell Sci.*, 121(11):1785–1791, 1 June 2008.

[29] Daniel J Müller, Jonne Helenius, David Alsteens, and Yves F Dufrêne. Force probing surfaces of living cells to molecular resolution. *Nat. Chem. Biol.*, 5(6):383–390, June 2009.

[30] Johannes Preiner, Noriyuki Kodera, Jilin Tang, Andreas Ebner, Mario Brameshuber, Dieter Blaas, Nicola Gelbmann, Hermann J Gruber, Toshio Ando, and Peter Hinterdorfer. IgGs are made for walking on bacterial and viral surfaces. *Nat. Commun.*, 5:4394, 10 July 2014.

[31] Yoshikazu Tsukasaki, Kazuo Kitamura, Kazuya Shimizu, Atsuko H Iwane, Yoshimi Takai, and Toshio Yanagida. Role of multiple bonds between the single cell adhesion molecules, nectin and cadherin, revealed by high sensitive force measurements. *J. Mol. Biol.*, 367(4): 996–1006, 6 April 2007.

[32] Linda Wildling, Christian Rankl, Thomas Haselgrübler, Hermann J Gruber, Marion Holy, Amy Hauck Newman, Mu-Fa Zou, Rong Zhu, Michael Freissmuth, Harald H Sitte, and Peter Hinterdorfer. Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells. *J. Biol. Chem.*, 287(1):105–113, 2 January 2012.

[33] Amy E M Beedle, Aisling Williams, Josep Relat-Goberna, and Sergi Garcia-Manyes. Mechanobiology-chemical origin of membrane mechanical resistance and force-dependent signaling. *Curr. Opin. Chem. Biol.*, 29(C):87–93, December 2015.

[34] Harald Janovjak, Jens Struckmeier, Maurice Hubain, Alexej Kedrov, Max Kessler, and Daniel J Müller. Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure*, 12(5):871–879, May 2004.

[35] Daniel J Müller. AFM: a nanotool in membrane biology. *Biochemistry*, 47(31):7986–7998, August 2008.

[36] Daniel J Müller and Andreas Engel. Atomic force microscopy and spectroscopy of native membrane proteins. *Nat. Protoc.*, 2(9):2191–2197, 2007.

[37] Amy E M Beedle, Ainhoa Lezamiz, Guillaume Stirnemann, and Sergi Garcia-Manyes. The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat. Commun.*, 6:7894, 3 August 2015.

[38] Tianjia Bu, Hui-Chuan Eileen Wang, and Hongbin Li. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. *Langmuir*, 28(33):12319–12325, 21 August 2012.

[39] Yi Cao, Yongnan Devin Li, and Hongbin Li. Enhancing the mechanical stability of proteins through a cocktail approach. *Biophys. J.*, 100(7):1794–1799, 6 April 2011.

[40] Armando del Rio, Raul Perez-Jimenez, Ruchuan Liu, Pere Roca-Cusachs, Julio M Fernandez, and Michael P Sheetz. Stretching single talin rod molecules activates vinculin binding. *Science*, 323(5914):638–641, 30 January 2009.

[41] Michael Geisler, Senbo Xiao, Elias M Puchner, Frauke Gräter, and Thorsten Hugel. Controlling the structure of proteins at surfaces. *J. Am. Chem. Soc.*, 132(48):17277–17281, 8 December 2010.

[42] Marta Kocun, Michel Grandbois, and Louis A Cuccia. Single molecule atomic force microscopy and force spectroscopy of chitosan. *Colloids Surf. B Biointerfaces*, 82(2): 470–476, 1 February 2011.

[43] M Rief, F Oesterhelt, B Heymann, and H E Gaub. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science (80-.).*, 275(February 1997): 1295–1297, 28 February 1997.

[44] Christian Albrecht, Kerstin Blank, Mio Lalic-Mülthaler, Siegfried Hirler, Thao Mai, Ilka Gilbert, Susanne Schiffmann, Tom Bayer, Hauke Clausen-Schaumann, and Hermann E Gaub. DNA: a programmable force sensor. *Science*, 301(5631):367–370, 18 July 2003.

[45] David Alsteens, Melissa C Garcia, Peter N Lipke, and Yves F Dufrêne. Force-induced formation and propagation of adhesion nanodomains in living fungal cells. *Proceedings of the National Academy of Sciences*, 107(48):20744–20749, 2010.

[46] Moritz Pfreundschuh, David Alsteens, Ralph Wieneke, Cheng Zhang, Shaun R Coughlin, Robert Tampé, Brian K Kobilka, and Daniel J Müller. Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM. *Nat. Commun.*, 6(8857):8857, 12 November 2015.

[47] Filomena A Carvalho, Ivo C Martins, and Nuno C Santos. Atomic force microscopy and force spectroscopy on the assessment of protein folding and functionality. *Arch. Biochem. Biophys.*, 531(1-2):116–127, March 2013.

[48] Ignacio Casuso, Félix Rico, and Simon Scheuring. Biological AFM: where we come from - where we are - where we may go. *J. Mol. Recognit.*, 24(3):406–413, May 2011.

[49] Toni Hoffmann and Lorna Dougan. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.*, 41(14):4781–4796, 21 July 2012.

[50] Chih-Kung Lee, Yu-Ming Wang, Long-Sun Huang, and Shiming Lin. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron*, 38(5):446–461, 2007.

[51] Hongbin Li and Yi Cao. Protein mechanics: from single molecules to functional biomaterials. *Acc. Chem. Res.*, 43(10):1331–1341, 19 October 2010.

[52] Piotr E Marszalek and Yves F Dufrêne. Stretching single polysaccharides and proteins using atomic force microscopy. *Chem. Soc. Rev.*, 41(9):3523–3534, 7 May 2012.

[53] Daniel J Müller and Yves F Dufrêne. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nat. Nanotechnol.*, 3(5):261–269, May 2008.

[54] Keir C Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*, 5(6):491–505, June 2008.

[55] Aleksandr Noy. Force spectroscopy 101: how to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment. *Curr. Opin. Chem. Biol.*, 15(5): 710–718, October 2011.

[56] Matthias Rief and Helmut Grubmüller. Force spectroscopy of single biomolecules. *Chemphyschem*, 3(3):255–261, 12 March 2002.

[57] Donald J Sirbuly, Raymond W Friddle, Joshua Villanueva, and Qian Huang. Nanomechanical force transducers for biomolecular and intracellular measurements: is there room to shrink and why do it? *Rep. Prog. Phys.*, 78(2):024101, February 2015.

[58] Michael T Woodside and Steven M Block. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu. Rev. Biophys.*, 43(1):19–39, May 2014.

[59] Joanna I Sulkowska and Marek Cieplak. Mechanical stretching of proteins—a theoretical survey of the protein data bank. *J. Phys. Condens. Matter*, 19(28):283201, 2007.

[60] E L Florin, V T Moy, and H E Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264(5157):415–417, 15 April 1994.

[61] V T Moy, E L Florin, and H E Gaub. Intermolecular forces and energies between ligands and receptors. *Science*, 266(5183):257–259, 14 October 1994.

[62] Félix Rico and Vincent T Moy. Energy landscape roughness of the streptavidin-biotin interaction. *J. Mol. Recognit.*, 20(6):495–501, November 2007.

[63] Chunbo Yuan, Aileen Chen, Pamela Kolb, and Vincent T Moy. Energy Landscape of Streptavidin-Biotin Complexes Measured by Atomic Force Microscopy. *Biochemistry*, 39 (33):10219–10223, August 2000.

[64] P Hinterdorfer, W Baumgartner, H J Gruber, K Schilcher, and H Schindler. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 93(8):3477–3481, 16 April 1996.

[65] Julia Morfill, Kerstin Blank, Christian Zahnd, Beatrice Luginbühl, Ferdinand Kühner, Kay-E Gottschalk, Andreas Plückthun, and Hermann E Gaub. Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. *Biophys. J.*, 93 (November):3583–3590, 15 November 2007.

[66] F Schwesinger, R Ros, T Strunz, D Anselmetti, H J Güntherodt, A Honegger, L Jermutus, L Tiefenauer, and A Pluckthun. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proc. Natl. Acad. Sci. U. S. A.*, 97(18): 9972–9977, 29 August 2000.

[67] Gabriel Mitchell, Charles-Antoine Lamontagne, Réjean Lebel, Michel Grandbois, and François Malouin. Single-molecule dynamic force spectroscopy of the fibronectin-heparin interaction. *Biochem. Biophys. Res. Commun.*, 364(3):595–600, 21 December 2007.

[68] Sebastian W Schmidt, Pavel Filippov, Alfred Kersch, Martin K Beyer, and Hauke Clausen-Schaumann. Single-molecule force-clamp experiments reveal kinetics of mechanically activated silyl ester hydrolysis. *ACS Nano*, 6(2):1314–1321, 28 February 2012.

[69] M A Jobst, L F Milles, C Schoeler, W Ott, D B Fried, E A Bayer, H E Gaub, and M A Nash. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLife*, 2015.

[70] Marcus Otten, Wolfgang Ott, Markus A Jobst, Lukas F Milles, Tobias Verdorfer, Diana A Pippig, Michael A Nash, and Hermann E Gaub. from genes to protein mechanics on a chip. *Nature Methods*, pages 1–6, September 2014.

[71] Markus A Jobst, Constantin Schoeler, Klara Malinowska, and Michael A Nash. Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy. *Journal of visualized experiments : JoVE*, (82), 2013.

[72] Constantin Schoeler, Rafael C Bernardi, Klara H Malinowska, Ellis Durner, Wolfgang Ott, Edward A Bayer, Klaus Schulten, Michael A Nash, and Hermann E Gaub. Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Letters*, 15(11): 7370–7376, November 2015.

[73] S W Stahl, M A Nash, D B Fried, M Slutzki, Y Barak, E A Bayer, and H E Gaub. Single-molecule dissection of the high-affinity cohesin–dockerin complex. *Proceedings of the National Academy of Sciences*, 109(50):20431–20436, 2012.

[74] Morten Bertz, Matthias Wilmanns, and Matthias Rief. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proceedings of the National Academy of Sciences*, 106(32):13307–13310, 2009.

[75] Claire Verbelen, Hermann J Gruber, and Yves F Dufrêne. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. *Journal of Molecular Recognition*, 20(6):490–494, November 2007.

[76] Joyce Wong, Ashutosh Chilkoti, and Vincent T Moy. Direct force measurements of the streptavidin-biotin interaction. *Biomol. Eng.*, 16(1-4):45–55, 31 December 1999.

[77] Jan Philipp Junker and Matthias Rief. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. *Proc. Natl. Acad. Sci. U. S. A.*, 106(34):14361–14366, 25 August 2009.

[78] Fabian Baumann, Magnus S Bauer, Lukas F Milles, Alexander Alexandrovich, Hermann E Gaub, and Diana A Pippig. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nature nanotechnology*, 11(1):89–94, January 2016.

[79] G Neuert, C Albrecht, E Pamir, and H E Gaub. Dynamic force spectroscopy of the digoxigenin-antibody complex. *FEBS Lett.*, 580(2):505–509, 23 January 2006.

[80] Constantin Schoeler, Tobias Verdorfer, Hermann E Gaub, and Michael A Nash. Biasing effects of receptor-ligand complexes on protein-unfolding statistics. *Physical Review E*, 94:042412, October 2016.

[81] Hendrik Dietz and Matthias Rief. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. U. S. A.*, 103(5):1244–1247, 31 January 2006.

[82] Diana A Pippig, Fabian Baumann, Mathias Strackharn, Daniela Aschenbrenner, and Hermann E Gaub. Protein-DNA chimeras for nano assembly. *ACS Nano*, 8(7):6551–6555, 22 July 2014.

[83] Jun Yin, Paul D Straight, Shaun M McLoughlin, Zhe Zhou, Alison J Lin, David E Golan, Neil L Kelleher, Roberto Kolter, and Christopher T Walsh. Genetically encoded short peptide tag for versatile protein labeling by sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U. S. A.*, 102(44):15815–15820, 1 November 2005.

[84] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by sfp phosphopantetheinyl transferase. *Nat. Protoc.*, 1(1):280–285, 2006.

[85] Julia L Zimmermann, Thomas Nicolaus, Gregor Neuert, and Kerstin Blank. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nature Protocols*, 5(6):975–985, May 2010.

[86] Mario B Viani, Tilman E Schäffer, Ami Chand, Matthias Rief, Hermann E Gaub, and Paul K Hansma. Small cantilevers for force spectroscopy of single molecules. *J. Appl. Phys.*, 86(4):2258–2262, 15 August 1999.

[87] M B Viani, T E Schäffer, G T Paloczi, L I Pietrasanta, B L Smith, J B Thompson, M Richter, M Rief, H E Gaub, K W Plaxco, A N Cleland, H G Hansma, and P K Hansma. Fast imaging and fast force spectroscopy of single biopolymers with a new atomic force microscope designed for small cantilevers. *Rev. Sci. Instrum.*, 70(11):4300–4303, 1 November 1999.

[88] Hao Yu, Derek R Dee, Xia Liu, Angela M Brigley, Iveta Sosova, and Michael T Woodside. Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. *Proceedings of the National Academy of Sciences*, 112(27):8308–8313, 7 July 2015.

[89] Gabriel Žoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 110(45):18156–18161, 5 November 2013.

[90] Devin T Edwards, Jaevyn K Faulk, Aric W Sanders, Matthew S Bull, Robert Walder, Marc-Andre LeBlanc, Marcelo C Sousa, and Thomas T Perkins. Optimizing 1μs-resolution Single-Molecule force spectroscopy on a commercial atomic force microscope. *Nano Lett.*, 15(10):7091–7098, October 2015.

[91] Hoi Sung Chung, Kevin McHale, John M Louis, and William A Eaton. Single-molecule fluorescence experiments determine protein folding transition path times. *Science*, 335 (6071):981–984, 24 February 2012.

[92] Benjamin Schuler and Hagen Hofmann. Single-molecule spectroscopy of protein folding dynamics–expanding scope and timescales. *Curr. Opin. Struct. Biol.*, 23(1):36–47, February 2013.

[93] Mingdong Dong and Ozgur Sahin. A nanomechanical interface to rapid single-molecule interactions. *Nat. Commun.*, 2(247):247, 2011.

[94] Chengzhi He, Georgi Z Genchev, Hui Lu, and Hongbin Li. Mechanically Untying a Protein Slipknot: Multiple Pathways Revealed by Force Spectroscopy and Steered Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 134(25):10428–10435, June 2012.

[95] Felix Rico, Laura Gonzalez, Ignacio Casuso, Manel Puig-Vidal, and Simon Scheuring. High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science*, 342(6159):741–743, 2013.

[96] Peter L Freddolino, Feng Liu, Martin Gruebele, and Klaus Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.*, 94(10):L75–7, 15 May 2008.

[97] B Heymann and H Grubmüller. Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces. *Biophys. J.*, 81(3):1295–1313, September 2001.

[98] Eric H Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery through the computational microscope. *Structure*, 17(10):1295–1306, 14 October 2009.

[99] Felix Rico, Laura Gonzalez, Ignacio Casuso, Manel Puig-Vidal, and Simon Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 8 November 2013.

[100] T Ando, N Kodera, E Takai, D Maruyama, K Saito, and A Toda. A high-speed atomic force microscope for studying biological macromolecules. *Proc. Natl. Acad. Sci. U. S. A.*, 98(22):12468–12472, 23 October 2001.

[101] Robert Walder, D Hern Paik, Matthew S Bull, Carl Sauer, and Thomas T Perkins. Ultra-stable measurement platform: sub-nm drift over hours in 3D at room temperature. *Opt. Express*, 23(13):16554–16564, 29 June 2015.

[102] Allison B Churnside, Ruby May A Sullan, Duc M Nguyen, Sara O Case, Matthew S Bull, Gavin M King, and Thomas T Perkins. Routine and timely sub-piconewton force stability and precision for biological applications of atomic force microscopy. *Nano Lett.*, 12(7): 3557–3561, 11 July 2012.

[103] Elio A Abbondanzieri, William J Greenleaf, Joshua W Shaevitz, Robert Landick, and Steven M Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–465, 24 November 2005.

[104] Jing Zhou, Volker Schweikhard, and Steven M Block. Single-molecule studies of RNAPII elongation. *Biochim. Biophys. Acta*, 1829(1):29–38, January 2013.

[105] Wei Cheng, Srikesh G Arunajadai, Jeffrey R Moffitt, Ignacio Tinoco, Jr, and Carlos Bustamante. Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. *Science*, 333(6050):1746–1749, 23 September 2011.

[106] Hao Yu, Xia Liu, Krishna Neupane, Amar Nath Gupta, Angela M Brigley, Allison Solanki, Iveta Sosova, and Michael T Woodside. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proceedings of the National Academy of Sciences*, 109(14):5283–5288, 3 April 2012.

[107] Devin T Edwards and Thomas T Perkins. Journal of Structural Biology. *Journal of structural biology*, pages 1–13, February 2016.

[108] Patrick D Bosshart, Patrick L T M Frederix, and Andreas Engel. Reference-free alignment and sorting of single-molecule force spectroscopy data. *Biophys. J.*, 102(9):2202–2211, 2 May 2012.

[109] M Kuhn, H Janovjak, M Hubain, and D J Müller. Automated alignment and pattern recognition of single-molecule force spectroscopy data. *J. Microsc.*, 218(Pt 2):125–132, May 2005.

[110] Elias M Puchner, Gereon Franzen, Mathias Gautel, and Hermann E Gaub. Comparing Proteins by Their Unfolding Pattern. *Biophysical Journal*, 95(1):426–434, July 2008.

[111] C Bustamante, J F Marko, E D Siggia, and S Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 9 September 1994.

[112] C Ortiz and G Hadziioannou. Entropic Elasticity of Single Polymer Chains of Poly(methacrylic acid) Measured by Atomic Force Microscopy. *Macromolecules*, 32 (3):780–787, 1999.

[113] L Livadaru, R R Netz, and H J Kreuzer. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules*, 36(10):3732–3744, May 2003.

[114] Johannes Thoma, Björn M Burmann, Sebastian Hiller, and Daniel J Müller. Impact of holdase chaperones Skp and SurA on the folding of beta-barrel outer-membrane proteins. *Nat. Struct. Mol. Biol.*, 22(10):795–802, September 2015.

[115] M Rief, M Gautel, F Oesterhelt, J M Fernandez, and H E Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 16 May 1997.

[116] A Valbuena, J Oroz, R Hervás, A M Vera, D Rodríguez, M Menéndez, J I Sulkowska, M Cieplak, and M Carrión-Vázquez. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences*, 106 (33):13791–13796, 2009.

[117] Thorsten Hugel, Matthias Rief, Markus Seitz, Hermann E Gaub, and Roland R Netz. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-Initio Theory. *Physical Review Letters*, 94(4):048301, January 2005.

[118] G Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, May 1978.

[119] E Evans and K Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72(4):1541–1555, April 1997.

[120] S Izrailev, S Stepaniants, M Balsera, Y Oono, and K Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal*, 72(4):1568–1581, April 1997.

[121] O K Dudko, G Hummer, and A Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, 105(41):15755, 2008.

[122] Raymond W Friddle, Aleksandr Noy, and James J De Yoreo. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proceedings of the National Academy of Sciences*, 109(34):13573–13578, August 2012.

[123] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Physical Review Letters*, 96(10): 108101, March 2006.

[124] H A Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.

[125] A Garg. Escape-Field Distribution for Escape From a Metastable Potential Well Subject to a Steadily Increasing Bias Field. *Physical Review B*, 51(21):15592–15595, 1995.

[126] A F Oberhauser, P K Hansma, M Carrión-Vázquez, and J M Fernandez. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proceedings of the National Academy of Sciences*, 98(2):468–472, January 2001.

[127] L Verlet. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103, 1967.

[128] A D MacKerell, D Bashford, Bellott, R L Dunbrack, J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiórkiewicz-Kuczera, D Yin, and M Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, April 1998.

[129] Eric H Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery Through the Computational Microscope. *Structure*, 17(10):1295–1306, October 2009.

[130] Guha Jayachandran, V Vishal, and Vijay S Pande. Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *The Journal of Chemical Physics*, 124(16):164902, 2006.

[131] Gongpu Zhao, Juan R Perilla, Ernest L Yufenyuy, Xin Meng, Bo Chen, Jiying Ning, Jinwoo Ahn, Angela M Gronenborn, Klaus Schulten, Christopher Aiken, and Peijun Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451):643–646, 2013.

[132] Shanmugapriya Sothiselvam, Bo Liu, Wei Han, Haripriya Ramu, Dorota Klepacki, Gemma Catherine Atkinson, Age Brauer, Maido Remm, Tanel Tenson, Klaus Schulten, Nora Vázquez-Laslop, and Alexander S Mankin. Macrolide antibiotics allosterically predispose the ribosome for translation arrest. *Proceedings of the National Academy of Sciences*, 111(27):9804–9809, July 2014.

[133] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. ScienceDirectMolecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, 31:64–74, April 2015.

[134] H Grubmüller, B Heymann, and P Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science*, 271(5251):997–999, 1996.

[135] Richard F Greene and Herbert B Callen. On the formalism of thermodynamic fluctuation theory. *Physical Review*, 83(6):1231, 1951.

[136] Burak Erman. Relationships between ligand binding sites, protein architecture and correlated paths of energy and conformational fluctuations. *Physical Biology*, 8(5):056003, August 2011.

[137] J C Phillips, R Braun, and W Wang. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 2005.

[138] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in tRNA: protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6620–6625, 2009.

[139] Roy H Doi and Akihiko Kosugi. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.*, 2(7):541–551, July 2004.

[140] A.L. Carvalho, F. Dias, J A M Prates, T. Nagy, H J Gilbert, Gideon J Davies, L. Ferreira, M.J. Romão, and C.M.G.A. Fontes. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. *Proc. Natl. Acad. Sci. U. S. A.*, 100(24):13809, 2003.

[141] Steven P Smith and Edward A Bayer. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.*, 23(5):686–694, October 2013.

[142] Edward A Bayer, Raphael Lamed, Bryan A White, and Harry J Flint. From cellulosomes to cellulosomics. *The Chemical Record*, 8(6):364–377, 2008.

[143] A L Demain, M Newcomb, and J H D Wu. Cellulase, Clostridia, and Ethanol. *Microbiology and Molecular Biology Reviews*, 69(1):124–154, March 2005.

[144] Sadanari Jindou, Jennifer M Brulc, Maly Levy-Assaraf, Marco T Rincon, Harry J Flint, Margret E Berg, Melissa K Wilson, Bryan A White, Edward A Bayer, Raphael Lamed, and Ilya Borovok. Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. *FEMS Microbiol. Lett.*, 285(2): 188–194, August 2008.

[145] S Y Ding, M T Rincon, R Lamed, J C Martin, S I McCrae, V Aurilia, Y Shoham, E A Bayer, and H J Flint. Cellulosomal Scaffoldin-Like Proteins from Ruminococcus flavefaciens. *Journal of bacteriology*, 183(6):1945–1953, March 2001.

[146] Marco T Rincon, Bareket Dassa, Harry J Flint, Anthony J Travis, Sadanari Jindou, Ilya Borovok, Raphael Lamed, Edward A Bayer, Bernard Henrissat, Pedro M Coutinho, Dion A Antonopoulos, Margret E Berg Miller, and Bryan A White. Abundance and Diversity of Dockerin-Containing Proteins in the Fiber-Degrading Rumen Bacterium, Ruminococcus flavefaciens FD-1. *PLoS One*, 5(8):e12476, August 2010.

[147] Michael E Himmel, Shi-You Ding, D K Johnson, William S Adney, Mark R Nimlos, J W Brady, and T D Foust. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, 315(5813):804–807, February 2007.

[148] Henri-Pierre Fierobe, Edward A Bayer, Chantal Tardif, Mirjam Czjzek, Adva Mechaly, Anne Bélaïch, Raphael Lamed, Yuval Shoham, and Jean-Pierre Bélaïch. Degradation of cellulose substrates by cellulosome chimeras Substrate targeting versus proximity of enzyme components. *J. Biol. Chem.*, 277(51):49621–49630, 2002.

[149] Alejandro Valbuena, Javier Oroz, Rubén Hervás, Andrés Manuel Vera, David Rodríguez, Margarita Menéndez, Joanna I Sulkowska, Marek Cieplak, and Mariano Carrión-Vázquez. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl. Acad. Sci. U. S. A.*, 106(33):13791–13796, August 2009.

[150] Jarrett J Adams, Bradley A Webb, Holly L Spencer, and Steven P Smith. Structural Characterization of Type II Dockerin Module from the Cellulosome of Clostridium thermocellum: Calcium-Induced Effects on Conformation and Target Recognition †. *Biochemistry*, 44(6):2173–2182, February 2005.

[151] Jarrett J Adams, Gour Pal, Zongchao Jia, and Steven P Smith. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U. S. A.*, 103(2):305–310, January 2006.

[152] Mateusz Sikora and Marek Cieplak. Mechanical stability of multidomain proteins and novel mechanical clamps. *Proteins: Structure, Function, and Bioinformatics*, 79(6):1786–1799, April 2011.

[153] Roman Brunecky, Markus Alahuhta, Yannick J Bomble, Qi Xu, John O Baker, Shi-You Ding, Michael E Himmel, and Vladimir V Lunin. Structure and function of the Clostridium thermocellum cellobiohydrolase A X1-module repeat: enhancement through stabilization of the CbhA complex. *Acta Crystallographica Section D-Biological Crystallography*, 68: 292–299, March 2012.

[154] Stefan W Stahl, Michael A Nash, Daniel B Fried, Michal Slutzki, Yoav Barak, Edward A Bayer, and Hermann E Gaub. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U. S. A.*, 109(50):20431–20436, December 2012.

[155] R Merkel, P Nassoy, A Leung, K Ritchie, and Evan Evans. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, 1999.

[156] Julia Morfill, Kerstin Blank, Christian Zahnd, Beatrice Luginbühl, Ferdinand Kühner, Kay-E Gottschalk, Andreas Plückthun, and Hermann E Gaub. Affinity-Matured Recombinant Antibody Fragments Analyzed by Single-Molecule Force Spectroscopy. *Biophysical Journal*, 93(10):3583–3590, November 2007.

[157] Felix Berkemeier, Morten Bertz, Senbo Xiao, Nikos Pinotsis, Matthias Wilmanns, Frauke Gräter, and Matthias Rief. Fast-folding α-helices as reversible strain absorbers in the muscle protein myomesin. *Proc. Natl. Acad. Sci. U. S. A.*, 108(34):14139–14144, 2011.

[158] Piotr E Marszalek, Hui Lu, Hongbin Li, Mariano Carrión-Vázquez, Andres F Oberhauser, Klaus Schulten, and Julio M Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–103, November 1999.

[159] Michel Grandbois, Martin Beyer, Matthias Rief, Hauke Clausen-Schaumann, and Hermann E Gaub. How Strong Is a Covalent Bond? *Science*, 283(5408):1727–1730, 1999.

[160] Yurui Xue, Xun Li, Hongbin Li, and Wenke Zhang. Quantifying thiol-gold interactions towards the efficient strength control. *Nature Communications*, 5, July 2014.

[161] Y J Bomble, G T Beckham, J F Matthews, M R Nimlos, M E Himmel, and M F Crowley. Modeling the self-assembly of the cellulosome enzyme complex. *Journal of Biological Chemistry*, 286(7):5614–5623, 2011.

[162] Marcos Sotomayor and Klaus Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316(5828):1144–1148, 2007.

[163] W Thomas, M Forero, O Yakovenko, L Nilsson, P Vicini, E Sokurenko, and V Vogel. Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophysical Journal*, 90(3):753–764, February 2006.

[164] W Wang and B A Malcolm. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques*, 26(4):680–682, 1999.

[165] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, April 2009.

[166] Asako Sawano and Atsushi Miyawaki. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic acids research*, 28(16):e78–e78, 2000.

[167] F William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, May 2005.

[168] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nature Protocols*, 1(1):280–285, June 2006.

[169] H Gumpp, S W Stahl, M Strackharn, E M Puchner, and H E Gaub. Ultrastable combined atomic force and total internal fluorescence microscope. *Review of Scientific Instruments*, 80(6):063704, 2009.

[170] Jeffrey L Hutter and John Bechhoefer. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64(7):1868, 1993.

[171] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[172] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *Journal of Computational Physics*, 151(1):283–312, May 1999.

[173] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro E M Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell, Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1and χ 2Dihedral Angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, September 2012.

[174] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926, 1983.

[175] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[176] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins Structure Function and Genetics*, 23(4):566–579, December 1995.

[177] João V Ribeiro, Juan A C Tamames, Nuno M F S A Cerqueira, Pedro A Fernandes, and Maria J Ramos. Volarea - A Bioinformatics Tool to Calculate the Surface Area and the Volume of Molecular Systems. *Chemical Biology & Drug Design*, 82(6):743–755, October 2013.

[178] R A Laskowski, E G Hutchinson, A D Michie, A C Wallace, M L Jones, and J M Thornton. PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences*, 22(12):488–490, December 1997.

[179] H Dietz and M Rief. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences*, 103(5):1244–1247, January 2006.

[180] G Bao and S Suresh. Cell and molecular mechanics of biological materials. *Nat. Mater.*, 2 (11):715–725, November 2003.

[181] A J Grodzinsky, M E Levenston, M Jin, and E H Frank. Cartilage tissue remodeling in response to mechanical forces. *Annu. Rev. Biomed. Eng.*, 2:691–713, 2000.

[182] Sanjay Kumar and Valerie M Weaver. Mechanics, malignancy, and metastasis: The force journey of a tumor cell. *Cancer Metast. Rev.*, 28(1-2):113–127, June 2009.

[183] V Vogel, W E Thomas, D W Craig, A Krammer, and G Baneyx. Structural insights into the mechanical regulation of molecular recognition sites. *Trends Biotechnol.*, 19(10):416–423, October 2001.

[184] Armando del Rio, Raul Perez-Jimenez, Ruchuan Liu, Pere Roca-Cusachs, Julio M Fernandez, and Michael P Sheetz. Stretching single talin rod molecules activates vinculin binding. *Science*, 323(5914):638–641, 2009.

[185] A Engel and D J Muller. Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Biol.*, 7(9):715–718, September 2000.

[186] Aleksandr Noy and Raymond W Friddle. Practical single molecule force spectroscopy: How to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope. *Methods*, 60(2):142–150, April 2013.

[187] James J De Yoreo, Sungwook Chung, and Raymond W Friddle. In Situ Atomic Force Microscopy as a Tool for Investigating Interactions and Assembly Dynamics in Biomolecular and Biomineral Systems. *Advanced Functional Materials*, 23(20):2525–2538, April 2013.

[188] Matthias Rief, Mathias Gautel, Alexander Schemmel, and Hermann E Gaub. The Mechanical Stability of Immunoglobulin and Fibronectin III Domains in the Muscle Protein Titin Measured by Atomic Force Microscopy. *Biophysical Journal*, 75(6):3008–3014, December 1998.

[189] Yongnan Devin Li, Guillaume Lamour, Jörg Gsponer, Peng Zheng, and Hongbin Li. The Molecular Mechanism Underlying Mechanical Anisotropy of theProtein GB1. *Biophys. J.*, 103(11):2361–2368, 2012.

[190] Whasil Lee, Xiancheng Zeng, Kristina Rotolo, Ming Yang, Christopher J Schofield, Vann Bennett, Weitao Yang, and Piotr E Marszalek. Mechanical Anisotropy of Ankyrin Repeats. *Biophys. J.*, 102(5):1118–1126, 2012.

[191] Bharat Jagannathan, Phillip J Elms, Carlos Bustamante, and Susan Marqusee. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc. Natl. Acad. Sci. U.S.A.*, 109(44):17820–17825, 2012.

[192] Ionel Popa, Ronen Berkovich, Jorge Alegre-Cebollada, Carmen L Badilla, Jaime Andrés Rivas-Pardo, Yukinori Taniguchi, Masaru Kawakami, and Julio M Fernandez. Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.*, 135(34):12762–12771, August 2013.

[193] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[194] R C Bernardi, M C R Melo, and K Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta*, 1850(5):872–877, 2015.

[195] A S T Ribeiro and V Ortiz. Energy Propagation and Network Energetic Coupling in Proteins. *J. Phys. Chem. B*, 119(5):1835–1846, 2015.

[196] A T Van Wart, J Durrant, L Votapka, and R E Amaro. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.*, 10(2):511–517, 2014.

[197] A T Van Wart, J Eargle, Z Luthey-Schulten, , and R E Amaro. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J. Chem. Theory Comput.*, 8(8):2949–2961, 2012.

[198] Rebecca W Alexander, John Eargle, and Zaida Luthey-Schulten. Experimental and computational determination of tRNA dynamics. *FEBS Lett.*, 584(2):376–386, January 2010.

[199] C Chennubhotla and I Bahar. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Syst. Biol.*, 2(36):1–13, 2006.

[200] Vesa P. Hytönen and Viola Vogel. A next-generation approach to the characterization of a non-model plant transcriptome. *PLoS Comput. Biol.*, 4(2):e24, 2008.

[201] Fatemeh Khalili-Araghi, James Gumbart, Po-Chao Wen, Marcos Sotomayor, Emad Tajkhorshid, and Klaus Schulten. Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.*, 19(2):128–37, 2009. ISSN 1879-033X.

[202] Weihua Li, Jie Shen, Guixia Liu, Yun Tang, and Tyuji Hoshino. Exploring coumarin egress channels in human cytochrome p450 2a6 by random acceleration and steered molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.*, 79(1):271–281, 2011.

[203] Wolfram Stacklies, M Cristina Vega, Matthias Wilmanns, and Frauke Gräter. Mechanical Network in Titin Immunoglobulin from Force Distribution Analysis. *PLoS Comput. Biol.*, 5(3):e1000306, March 2009.

[204] Senbo Xiao, Wolfram Stacklies, Murat Cetinkaya, Bernd Markert, and Frauke Graeter. Mechanical Response of Silk Crystalline Units from Force-Distribution Analysis. *Biophysical Journal*, 96(10):3997–4005, 2009.

[205] Wolfram Stacklies, Christian Seifert, and Frauke Graeter. Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinf.*, 12(1):101, 2011.

[206] Christian Seifert and Frauke Gräter. Force Distribution Reveals Signal Transduction in E. coli Hsp90. *Biophys. J.*, 103(10):2195–2202, 2012.

[207] Zoltan Palmai, Christian Seifert, Frauke Gräter, and Erika Balog. An Allosteric Signaling Pathway of Human 3-Phosphoglycerate Kinase from Force Distribution Analysis. *PLoS Comput. Biol.*, 10(1), 2014.

[208] Micha l Wojciechowski, Damien Thompson, and Marek Cieplak. Mechanostability of cohesin-dockerin complexes in a structure-based model: Anisotropy and lack of universality in the force profiles. *The Journal of Chemical Physics*, 141(24):245103, December 2014.

[209] Rommie E Amaro, Anurag Sethi, Rebecca S Myers, V Jo Davisson, and Zaida A Luthey-Schulten. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry*, 46(8):2156–2173, February 2007.

[210] J Eargle and Z A Luthey-Schulten. NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics*, 28(22):3000–3001, 2012.

[211] N Glykos. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.*, 27(14):1765–1768, 2006.

[212] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85(1):5–15, July 2003.

[213] O K Dudko, G Hummer, and A Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *PNAS*, 105(41):15755–15760, October 2008.

[214] M Karplus and J N Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.

[215] R M Levy, M Karplus, J Kushick, and D Perahia. Evaluation of the Configurational entropy for Proteins: Application to Molecular Dynamics Simulations of an α-Helix. *Macromolecules*, 17(7):1370–1374, 1984.

[216] Rafael C Bernardi, Isaac Cann, and Klaus Schulten. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels*, 7(83):1–8, 2014.

[217] Oliver F Lange and Helmut J Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, March 2006.

[218] Jaime Andrés Rivas-Pardo, Edward C Eckels, Ionel Popa, Pallav Kosuri, Wolfgang A Linke, and Julio M Fernandez. Work Done by Titin Protein Folding Assists Muscle Contraction. *CellReports*, pages 1–10, February 2016.

[219] Armando del Rio, Raul Perez-Jimenez, Ruchuan Liu, Pere Roca-Cusachs, Julio M Fernandez, and Michael P Sheetz. Stretching Single Talin Rod Molecules Activates Vinculin Binding. *Science*, 323(5914):638–641, 2009.

[220] Keir C Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5(6):491–505, June 2008.

[221] S B Smith, L Finzi, and C Bustamante. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 1992.

[222] O K Dudko. Decoding the mechanical fingerprints of biomolecules. *Quarterly reviews of biophysics*, 2015.

[223] B Heymann and H Grubmüller. Dynamic force spectroscopy of molecular adhesion bonds. *Physical Review Letters*, 84(26):6126–6129, 2000.

[224] Jakob T Bullerjahn, Sebastian Sturm, and Klaus Kroy. Theory of rapid force spectroscopy. *Nature Communications*, 5:1–10, July 2014.

[225] Raymond W Friddle. Unified Model of Dynamic Forced Barrier Crossing in Single Molecules. *Physical Review Letters*, 100(13):138302, April 2008.

[226] E Evans, A Leung, V Heinrich, and C Zhu. Mechanical switching and coupling between two dissociation pathways in a P-selectin adhesion bond. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31):11281–11286, 2004.

[227] C Friedsam, A K Wehle, F Kühner, and H E Gaub. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *Journal of Physics: Condensed Matter*, 15(18):S1709, 2003.

[228] S K Kufer, E M Puchner, H Gumpp, T Liedl, and H E Gaub. Single-Molecule Cut-and-Paste Surface Assembly. *Science*, 319(5863):594–596, February 2008.

List of Figures

List of Tables

Acknowledgments

I would like to thank a number of people without whom this thesis would not have been possible. First and foremost, I would like to thank my advisor **Hermann Gaub**. He has done an amazing job supervising me throughout my scientific career. It started with my Bachelor's thesis, where he made sure there was absolutely no friction between the university and attocube. During my Master's and time as a PhD student, he always managed to let me work independently and yet provide invaluable insight and ideas during discussions. I have particularly enjoyed our numerous discussions about force propagation and correlation analysis.

Second, I would like to thank **Michael Nash**. Michael did a great job of putting together an awesome team - the infamous "NashCats". He really helped to get me off to a great start back in 2012 with lots of valuable input. He was always available to discuss virtually anything and writing papers with him has always been a pleasure. Looking forward to seeing you in action in Basel, Professor Nash!

None of the results that lead to this thesis would have been possible without the members of the NashCats and quite frankly, time around the office would have been somewhat boring! My special thanks go to: **Ellis Durner** for being able to fix basically anything ("Ellis! Löten!"), his passion for building and tinkering with instruments, and putting up with the rest of us when we make fun of you for procrastinating. You should have kept the house though. **Markus Jobst** for being an amazing colleague from day one and a great discussion partner, your endless efforts in developing and improving the Schnecke, and your passion for good science. **Klara Malinowska** for the great time we had both during our projects in the lab, her passion for detail and for simply being a very good friend. **Lukas Milles** for the countless hours he invested in the analysis software, without which analyzing our experiments would be a nightmare, his appreciation of solid research, and inspiring theoretical discussions. **Wolfgang Ott** for his constructive skepticism, unbelievable work ethic, and most importantly for his immense knowledge on molecular biology. We would still be in the Stone Age without you. **Tobias Verdorfer**, who has been one of my closest friends throughout our studies, for your wildly inappropriate sense of humor, appreciation for well built stuff, and countless fun memories.

I would further like to thank the entire **Gambicrew** for making this group such an amazing place to work in. Specially, I want to thank **Fabian Baumann** and **Magnus Bauer** for lots of discussions on physics and one or two matches at the Kicker. While the Kicker should get his own acknowledgments section, I will resort to thanking the people who helped operating our most beloved scattering experiment: **Achim Löf, Philipp Walker, and Leonard Schendel**.

Thanks to **Diana Pippig** for assistance with many biochemical questions, her willingness to engage in teasing matches, and for never quite giving up on our plan to go "eskalieren".

Also I would like to thank **Linda Brützel** for being a wonderful friend and lunch companion, always providing sanity breaks when needed, and the many amazing times we had.

While I cannot thank **Franziska Kriegel** enough for taking the time to carefully proofread my thesis and motivating me during the writing phase, she actually deserves credit that goes far beyond being a good co-worker. You are one of the most candid, genuine and funniest

people I know. Thank you for being such a great friend.

Most of the results that led to this thesis would not have been possible without the amazing collaboration with **Rafael Bernardi** at the Beckman Institute. Working with Rafael has been amazingly productive, very insightful and incredibly fun. He could not have been a better host during my visits to Urbana (*aka*, the Cornfields). I am happy that we have become such good friends over the years.

Last but not least I want to thank the late **Klaus Schulten**, to whom I want to dedicate this thesis. Not only did Klaus help set up the collaboration between Munich and Urbana that led to the most important findings of this thesis, but he also became a great mentor and true inspiration for me. Klaus was one of the most brilliant people I have ever met and unbelievably friendly and generous at the same time. His scientific heritage will continue to provide answers to fundamental questions in biophysics and our understanding of biology at the single molecule level.